

Article

Internet Traffic Prediction with Distributed Multi-Agent Learning

Weiwei Jiang ^{1,*} , Miao He ² and Weixi Gu ^{3,*}¹ School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China² Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China³ China Academy of Industrial Internet, Beijing 100102, China

* Correspondence: jww@bupt.edu.cn (W.J.); guweixi@china-caii.com (W.G.)

Abstract: Internet traffic prediction has been considered a research topic and the basis for intelligent network management and planning, e.g., elastic network service provision and content delivery optimization. Various methods have been proposed in the literature for Internet traffic prediction, including statistical, machine learning and deep learning methods. However, most of the existing approaches are trained and deployed in a centralized approach, without considering the realistic scenario in which multiple parties are concerned about the prediction process and the prediction model can be trained in a distributed approach. In this study, a distributed multi-agent learning framework is proposed to fill the research gap and predict Internet traffic in a distributed approach, in which each agent trains a base prediction model and the individual models are further aggregated with the cooperative interaction process. In the numerical experiments, two sophisticated deep learning models are chosen as the base prediction model, namely, long short-term memory (LSTM) and gated recurrent unit (GRU). The numerical experiments demonstrate that the GRU model trained with five agents achieves state-of-the-art performance on a real-world Internet traffic dataset collected in a campus backbone network in terms of root mean square error (RMSE) and mean absolute error (MAE).



Citation: Weiwei J.; Miao, H.; Weixi, G. Internet Traffic Prediction with Distributed Multi-Agent Learning.

Appl. Syst. Innov. **2022**, *5*, 121.

<https://doi.org/10.3390/asi5060121>

Academic Editor: Jay Lee

Received: 17 November 2022

Accepted: 25 November 2022

Published: 29 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: internet traffic prediction; distributed multi-agent learning; long short-term memory; gated recurrent unit

1. Introduction

Internet traffic has grown considerably in the past few years, with the development of new networking paradigms, including 5G/6G, the Internet of Things and the Industrial Internet, and new Internet applications, including live streaming, video sharing and virtual reality. A precise prediction for Internet traffic has been proposed as an important research topic in the past few years as the basis for intelligent network management and planning, e.g., elastic network service provision and content delivery optimization. Different approaches have been proposed for achieving satisfactory traffic prediction performance, which can be categorized into statistical, machine learning and deep learning types.

Statistical models mainly include autoregressive integrated moving average (ARIMA) and vector autoregression (VAR) when the Internet traffic prediction problem is modeled as a univariate or multivariate time series problem. These statistical models have the advantages of low computational cost and high interpretability. However, their prediction performance is not as competitive as that of machine learning and deep learning models [1]. Machine learning models mainly include support vector machine and decision tree-based models, e.g., random forest and AdaBoost, and have both a medium computation cost and a medium prediction performance. More recently, deep learning models including various neural network structures have been proven more effective for prediction problems [2–5]. With a flexible structure, deep learning can be used in different problem formats, e.g.,

recurrent neural networks for time series-format prediction [6,7], convolutional neural networks for grid-based prediction [8,9] and graph neural networks for graph-format prediction [10,11].

However, most of the existing studies for Internet traffic prediction focus on the centralized approach, in which the prediction models are trained and deployed in a centralized server, neglecting the case that multiple parties are concerned about the prediction process and want to be involved in the training process in a distributed approach. In reality, multiple parties would be interested in the prediction result of Internet traffic usages. For example, for the campus backbone network considered in this paper, the involved parties include the university network management department, the Internet service provider and the Internet content provider. All these parties behave as an intelligent agent with their own computation and communication facilities that can be leveraged in the traffic prediction task. On the other hand, distributed multi-agent learning has been proven more efficient than training a deep learning model on a single agent [12]. However, few previous discussed whether better performance can be achieved with such settings in the problem of Internet traffic prediction.

To the best of our knowledge, this paper presents a pioneer work which predicts Internet traffic with a distributed multi-agent learning approach when Internet traffic prediction is modeled as a supervised learning problem. For this paper, the research motivation is to validate the marginal benefit of the distributed multi-agent learning approach over traditional centralized learning and initiate the research direction of a potential further performance improvement in follow-up studies. Two recurrent neural networks, namely, long short-term memory (LSTM) [13] and gated recurrent unit (GRU) [14], are adopted as the base prediction model and compared on a real-world Internet traffic dataset. The numerical experiments show that the GRU-based distributed multi-agent learning approach achieves state-of-the-art performance, outperforming both the LSTM-based multi-agent learning approach and the centralized learning approach in a previous study [15].

The contributions of this paper are summarized as follows:

1. To the best of our knowledge, this paper presents a pioneer work which predicts Internet traffic with a distributed multi-agent learning approach when Internet traffic prediction is modeled as a supervised learning problem and the base prediction models are trained cooperatively among different agents.
2. An effective interaction process is used for coordinating the different agents in the distributed training step, which can be modeled and analyzed as an irreducible aperiodic Markov chain with a finite state, and the convergence property of the interaction process is proved.
3. The effectiveness of the proposed approach is validated with a real-world Internet traffic dataset collected at the State University of Ceará for half a year from 16 January 2019 to 15 July 2019, and the five-agent GRU-based distributed multi-agent learning scheme achieves state-of-the-art performance with the smallest prediction errors and outperforms several sophisticated deep learning models in terms of root mean square error (RMSE) and mean absolute error (MAE).

The remainder of this paper is organized as follows. In Section 2, different types of prediction models are further discussed. In Section 3, a real-world Internet traffic dataset is introduced and the prediction problem is defined. In Section 4, the distributed multi-agent learning approach as well as the LSTM and GRU models are introduced. In Section 5, numerical experiments are conducted with the traffic dataset and the results are analyzed. In Section 6, the conclusion and future research directions are discussed.

2. Related Work

In this section, a brief introduction for the different types of prediction models is given, including statistical, machine learning and deep learning models. More comprehensive discussions can be found in recent relevant surveys [16–18].

2.1. Statistical Prediction Models

Statistical models are featured by their explicit mathematical formulas based on a set of statistical assumptions for the data generation or distribution patterns. Their prediction performance can be analyzed and quantified in theory. Some representative statistical prediction models include ARIMA and Holt–Winters (HW) models. Other models include the simple exponential smoothing model, which is used for cellular traffic prediction [19], and the particle filtering method, which is used for 5G traffic demand prediction [20]. However, statistical prediction models are less common in recent years for network traffic prediction problems, because their prediction performance is worse than machine learning and deep learning models.

2.2. Machine Learning-Based Prediction Models

Machine learning models are applied for network traffic prediction because of their strong learning ability for mapping historical network traffic data to future situations in a supervised learning approach. With the accumulation of network traffic datasets and the growth of computation capacities, machine learning models are becoming popular in the computer network domain, not only for traffic prediction but also for a wider range of applications, e.g., traffic classification, routing and intrusion detection [21].

Compared with deep learning models represented by deep neural networks, the machine learning models discussed in this subsection are those with shallow structures, e.g., decision trees and random forests. These tree-based machine learning models represent better interpretability and less overfitting severity compared with neural networks and are still used in network traffic prediction tasks, e.g., Gaussian process regression has been adopted for network traffic prediction in different network scenarios [22–24].

Besides, different machine learning models can be used jointly, e.g., random forest (RF) and LightGBM are used together for mobile network traffic prediction when RF is used for feature filtering and LightGBM is used to make predictions [25]. Machine learning models can also be used together with statistical models for further performance improvement, e.g., the combination of soft clustering and traditional time series models [26].

2.3. Deep Learning-Based Prediction Models

As a subset of machine learning models, deep learning models are becoming the mainstream solutions for network traffic prediction problems in various scenarios due to their superior performance. Deep learning-based prediction models can be roughly divided into different categories based on the neuron connection method in neural networks, i.e., recurrent neural networks, convolutional neural networks and graph neural networks.

Recurrent neural networks (RNNs) have been proposed for handling long sequence data, e.g., natural languages and time series, by adding the recurrent connection and memory mechanism. LSTM and GRU are two representative RNN variants and solve the problem of vanishing gradients faced by the original RNN structure. They have been adopted in network traffic problems with both univariate and multivariate time series forecasting problem formulations [7,27,28].

When considering network traffic prediction problems in a regular grid, convolutional neural networks (CNNs) can be used by taking traffic data in different grids as image pixel values [29–31]. The attention mechanism [32] can be further incorporated into CNNs for a better prediction performance [33]. For example, the attention mechanism is combined with ConvLSTM for capturing long-term spatial-temporal dependency for cellular traffic prediction and helps to achieve accurate prediction under hourly and daily time scales [34]. Based on the Transformer network, a spatial-temporal downsampling neural network model is further proposed for citywide mobile traffic prediction [35].

More recently, graph neural networks (GNNs) have been proposed as the new frontier of artificial intelligence research, and the input data are formulated as graphs [36]. GNNs have been successfully applied in communication networks, with a wide range of applications including network traffic prediction [37–39]. To further improve the per-

formance of the GNN-based prediction model in large-scale traffic prediction, transfer learning is introduced for knowledge reuse and computation reduction in cellular traffic prediction [40].

While there are various network traffic prediction models in the literature, most of them are considered in a centralized approach. Thus, the main research motivation of this paper is to fill in the research gap that the distributed multi-agent learning approach has not been considered in the Internet traffic prediction problem before, in the cooperative way more specifically. Two main research questions arise in this paper, i.e., would the interaction among different agents converge in the long run, and could the distributed multi-distributed learning approach outperform the traditional centralized learning approach? From our research, both questions are answered in the affirmative.

3. Dataset and Problem

In this paper, we use an open real-world dataset [41] to validate the proposed distributed multi-agent learning approach and compare it with the centralized learning baseline. The dataset contains the Internet traffic data collected in a university campus backbone network located in the State University of Ceará for half a year from 16 January 2019 to 15 July 2019, with a sampling time period of one hour. The mean, maximum, and minimum bandwidth usage values are recorded. Figure 1 shows the Internet traffic data in the first three weeks of the data collection time period. For clarity, the mean, maximum, and minimum values are plotted in three subfigures. Both a weekly pattern and a daily pattern on weekdays are observed from Figure 1. The stationarity of this specific Internet traffic time series has also been validated with the Dickey-Fuller test, and the rationality of using the mean, maximum, and minimum usage values as input has also been justified with the cointegration test [15].

Following previous studies [15], the Internet traffic prediction problem is modeled as a supervised learning problem, in which the mean traffic value in the future is the short-term prediction target and the historical traffic data (containing the mean, maximum, and minimum usage values) in the historical time window are the input features. We consider the single-step prediction problem in this study, in which the mean traffic value in the next hour is chosen as the prediction target. To compare with existing centralized learning results, we choose 24 h as the historical time window size and leave the exploration for multi-step predictions as well as the influence of varying historical time window sizes in our future studies.

In this study, we take a step further and consider Internet traffic prediction using supervised learning and distributed multi-agent learning approach. The problem is to design an effective cooperation and aggregation process in which the historical traffic data are used by different agents as the base prediction model input and a final result is aggregated from the predictions of different agents as the output. Different from the centralized learning approach in which only one prediction model is trained in a central server, the learning approach for the base prediction models is distributed among different agents with their own facilities.

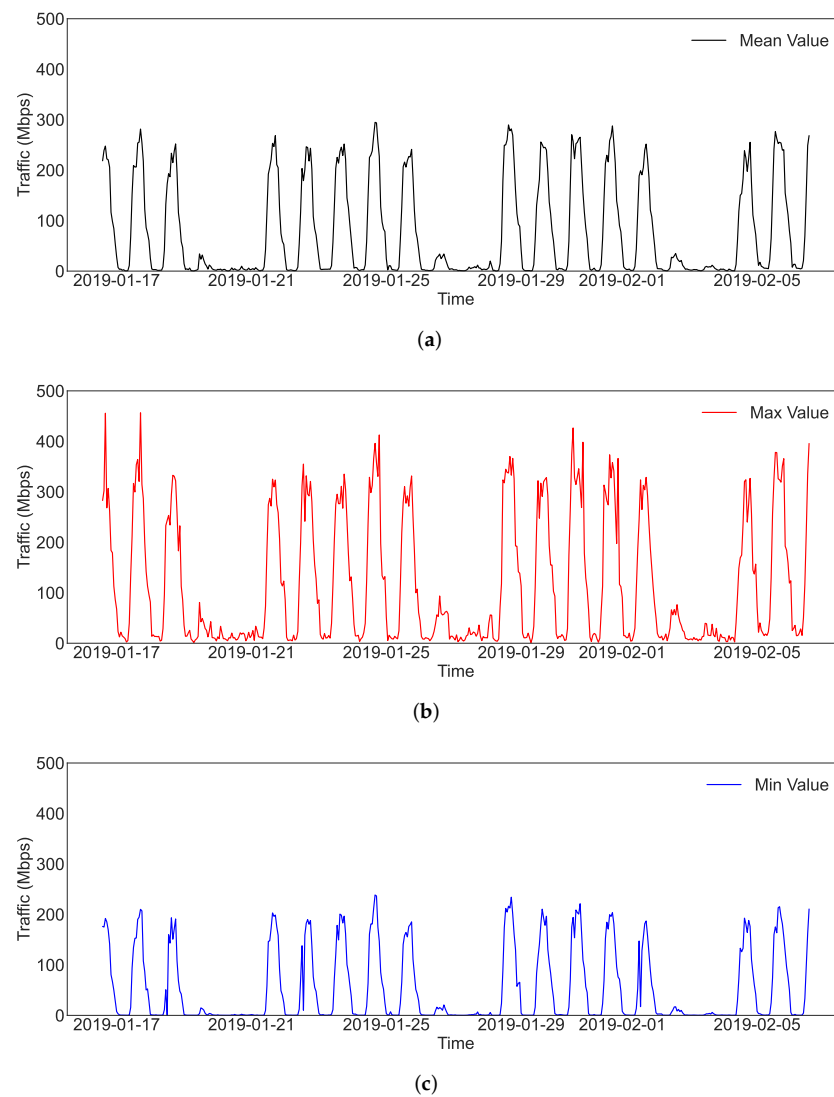


Figure 1. The Internet traffic data in the first three weeks. (a) Mean values; (b) Maximum values; (c) Minimum Values.

4. Methodology

In this paper, a distributed multi-agent learning approach is proposed for Internet traffic prediction, as shown in Figure 2. An interaction process is added into the individual training process so that the base prediction models are trained in a cooperative approach among the agents and an aggregated prediction is used as the final output, e.g., an average value. In this study, the base prediction model is considered a deep learning model with multiple layers and the described interaction is conducted at a layer-to-layer level for agents so that the prediction model can be cooperatively trained.

Mathematically, the influence between different agents in the interaction process is modeled as an adjacency matrix W with row sum 1. Denote that there are m agents and the agent set is I . Each agent trains a base prediction model with L layers. Denote $f_i^{(\ell)}$ as the mapping function for layer ℓ of agent i , to extract information from the input historical data. The distributed multi-agent learning in a single iteration is formulated as follows:

$$f_i^{(\ell+1)} = \sum_{j=1}^m w_{ij} f_j^{(\ell)} \quad (1)$$

where w_{ij} is an element of W , which is usually a positive value between 0 and 1. w_{ij} reflects the interaction degree among agent i and agent j . The above process can be iterated n times, and the convergence of the distributed multi-agent learning approach can be guaranteed when n is large enough [12]. The interaction among different agents is modeled as an irreducible aperiodic Markov chain with a finite state, in which a single stationary distribution $\{\pi_j, j \in I\}$ exists. Theoretically, when n goes to ∞ , the influence among agents converges to the stationary distribution, i.e., $\lim_{n \rightarrow \infty} w_{ij}^{(n)} = \pi_j, j \in I$. This observation answers our first research question that the interaction among different agents would converge in the long run in our proposed approach.

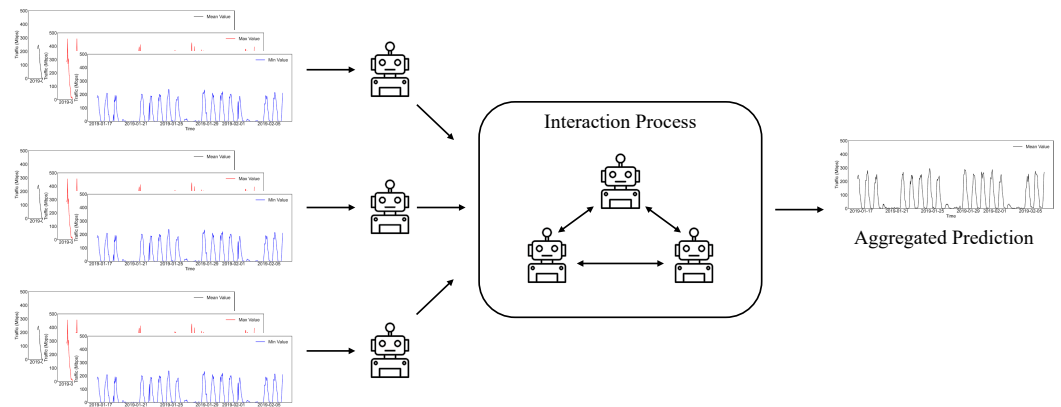


Figure 2. The general process of the proposed distributed multi-agent learning approach.

In the proposed approach, all the agents adopt the homogeneous base prediction model, e.g., LSTM or GRU, so that the computation requirement is the same and fair for different agents who want to be involved. An asynchronous communication requirement among the involved agents is also assumed in the proposed approach, so that the mappings can be exchanged and aggregated. It would be interesting to extend our study from the homogeneous and synchronous case to the heterogeneous or asynchronous cases when different agents have different computation and communication resources. For example, deep learning models with a higher computational burden can be trained on a computation-rich agent.

Two deep learning models are adopted in this paper, namely, LSTM and GRU. Based on the gate mechanism, LSTM is proposed to learn the long dependency relationship from the input sequence while mitigating the problem of vanishing gradients. GRU is a simplified variant of LSTM with only two gates, instead of three gates as LSTM, as shown in Figure 3. More specifically, denote t as the time point, x_t as the input, h_t as the output value, and c_t as the cell state. The three gates used in an LSTM cell can be denoted as follows. The forget gate is as follows:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2)$$

The input gate is as follows:

$$c_t = f_t * c_{t-1} + \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) * \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

And the output gate is as follows:

$$h_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) * \tanh(c_t) \quad (4)$$

where W_* and b_* are learnable parameters, $\sigma(*)$ is the sigmoid activation function and $\tanh(*)$ is the tanh activation function.

The two gates used in a GRU cell can be denoted similarly. The update gate is as follows:

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \quad (5)$$

And the output gate is as follows:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tanh(W_{xo}x_t + W_{ho}\sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) * h_{t-1} + b_o) \quad (6)$$

Two widely used evaluation metrics are adopted in this paper, namely, RMSE and MAE. Given the true values y and the predictions $\hat{y}$, RMSE is defined as

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (7)$$

and MAE is defined as

$$\text{MAE}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (8)$$

where N is the test subset size and the evaluation is conducted on the test subset.

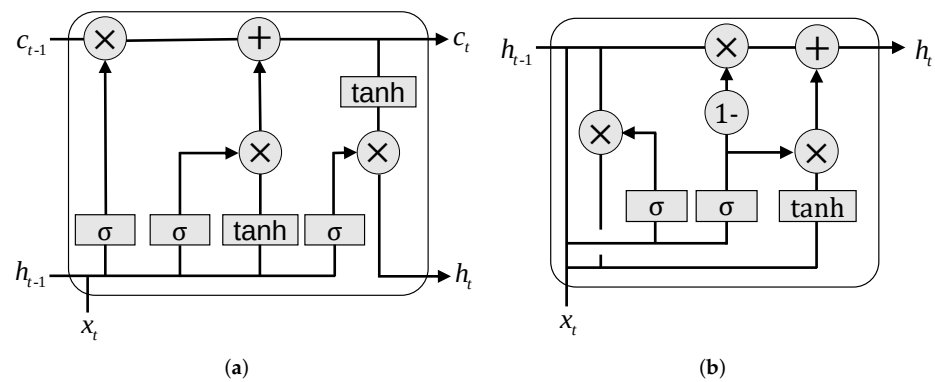


Figure 3. The LSTM and GRU cells. (a) LSTM; (b) GRU.

5. Experiment and Analysis

The Internet traffic dataset is divided into training and test subsets at a ratio of 5:1. The deep learning models are implemented with Python and TensorFlow, with three recurrent layers and 20 neurons in each recurrent layer. Adam is used as the optimizer with an adaptive learning rate starting from 0.01 and mean square error (MSE) is used as the loss function. The batch size is 128 and each model is trained for 200 epochs. For each agent number, ten runs are conducted separately. The weight matrix W for different agent numbers from the previous study [12] is used in this paper:

$$W_2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.15 & 0.85 \end{bmatrix} \quad (9)$$

$$W_3 = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.1 & 0.8 & 0.1 \\ 0.025 & 0.025 & 0.95 \end{bmatrix} \quad (10)$$

$$W_4 = \begin{bmatrix} 0.8 & 0.025 & 0.025 & 0.1 \\ 0.05 & 0.85 & 0.025 & 0.075 \\ 0.025 & 0.025 & 0.9 & 0.05 \\ 0.05 & 0.1 & 0.1 & 0.75 \end{bmatrix} \quad (11)$$

$$W_5 = \begin{bmatrix} 0.8 & 0.05 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.85 & 0.025 & 0.025 & 0.05 \\ 0.025 & 0.025 & 0.85 & 0.05 & 0.05 \\ 0.05 & 0.1 & 0.05 & 0.7 & 0.1 \\ 0.04 & 0.11 & 0.05 & 0.05 & 0.75 \end{bmatrix} \quad (12)$$

The evaluation results with different agent numbers for the LSTM and GRU models are shown in Figure 4, which are box and whisker plots for ten separate runs. With the increase in agent numbers, both LSTM and GRU demonstrate better performance with decreased RMSE and MAE values. GRU maintains a better performance than LSTM in Figure 4. While adding more agents into training might bring a further potential performance improvement, the marginal improvement becomes increasingly smaller. In addition, it becomes more difficult to coordinate more agents in practice. Thus, an agent number up to 5 is used in this paper. Since the research motivation of this paper is to validate the superiority of the distributed multi-agent learning approach over traditional centralized learning, we leave the exploration of using more agents in the follow-up studies.

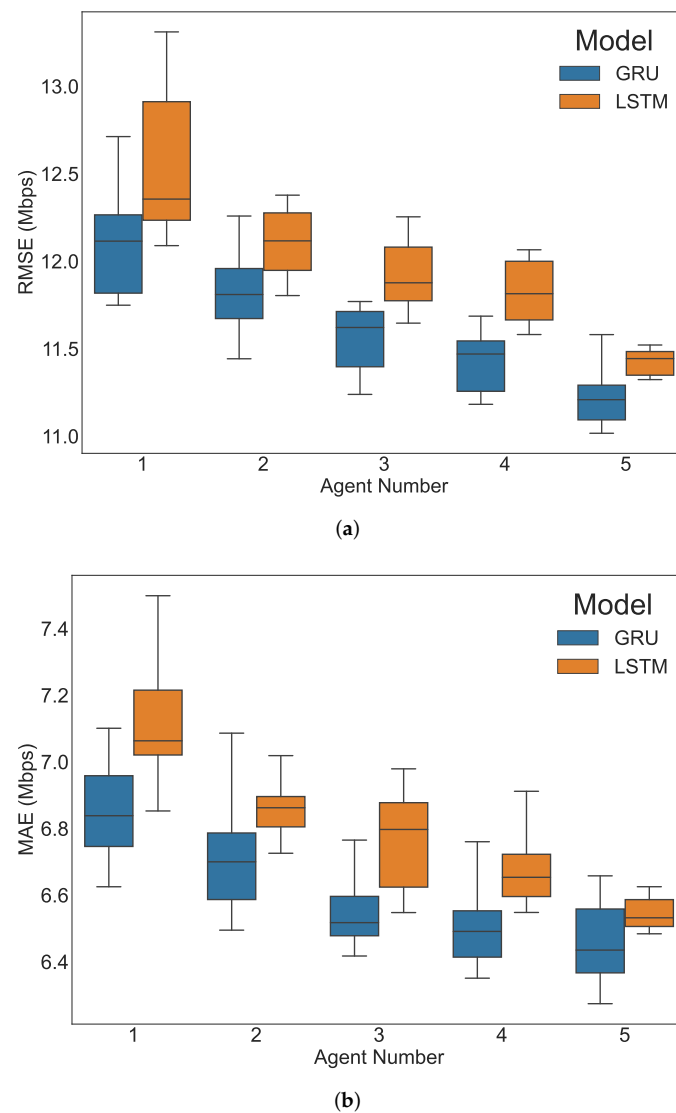


Figure 4. The evaluation results with different agent numbers. (a) RMSE; (b) MAE.

The mean RMSE and MAE results are further listed in Table 1 for LSTM and GRU models trained with different agent numbers. The baselines for comparison are those so-

phisticated deep learning models from previous studies, including Time Series Transformer (TST) [42], Multilayer Perceptron (MLP) [43], Temporal Convolutional Networks (TCN) [44], Fully Convolutional Network (FCN) [45], Residual Neural Network (ResNet) [45], and InceptionTime [46] in the traditional centralized learning approach. Among these baseline methods, the previous best prediction result is obtained with InceptionTime. While LSTM or GRU trained with a single agent is not competitive with InceptionTime, an improved performance is obtained by adding and involving more agents. Only three agents are enough for defeating InceptionTime when GRU is the base prediction model in individual agents, as indicated in Table 1. A new state-of-the-art performance in the literature is achieved for the considered Internet traffic data by GRU-based distributed multi-agent learning trained with five agents, as shown in bold in Table 1. The results in Table 1 answer our second research question that the proposed distributed multi-distributed learning approach manages to outperform the traditional centralized learning approach and would be worthy a further investigation in the future studies for similar problems.

Table 1. The experimental results.

Model	Agent Number	RMSE (Mbps)	MAE (Mbps)
TST [42]	N/A	16.28	9.71
MLP [43]	N/A	14.81	8.51
TCN [44]	N/A	13.39	9.31
FCN [45]	N/A	12.79	8.10
ResNet [45]	N/A	11.68	7.16
InceptionTime [46]	N/A	11.65	7.07
LSTM	1	12.57	7.10
	2	12.11	6.86
	3	11.93	6.77
	4	11.83	6.69
	5	11.43	6.54
GRU	1	12.11	6.84
	2	11.84	6.71
	3	11.56	6.55
	4	11.42	6.50
	5	11.22	6.45

Figure 5 further shows the comparison between true and predicted results for the first week in the test set. For clarity, only a single run with five agents is plotted for LSTM or GRU. From Figure 5, the predictions from both models present a similar pattern with the true values. However, LSTM generates larger predicted values when the true values are small, e.g., from 22 June 2019 to 23 June 2019, which results in a larger error than GRU. In other words, the GRU-based scheme is more stable, with less extreme predicted values than LSTM and a better overall prediction performance.

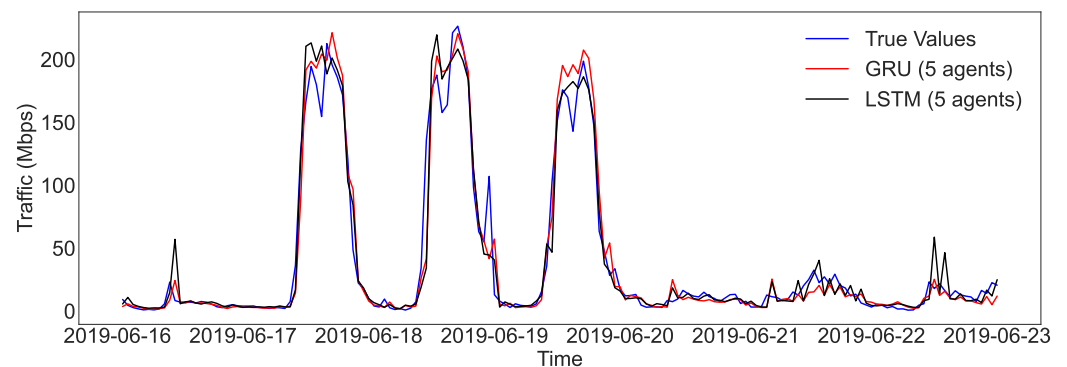


Figure 5. The comparison between true and predicted results for the first week in the test set.

6. Conclusions

The Internet traffic prediction problem is considered in this paper when multiple parties are involved, and a distributed multi-agent learning approach is proposed as the solution for the first time. In the proposed approach, an effective interaction process is used for coordinating the different agents in the distributed training step and the aggregated prediction from different agents is the final result. The interaction process is modeled and analyzed as an irreducible aperiodic Markov chain with a finite state, and its convergence is proven to be guaranteed. Numerical experiments are further conducted with LSTM or GRU as the prediction model, based on a real-world Internet traffic dataset collected at the State University of Ceará for half a year from 16 January 2019 to 15 July 2019. Compared with sophisticated deep learning baseline models trained in the centralized learning approach, the five-agent GRU-based distributed multi-agent learning scheme achieves a new state-of-the-art result, with RMSE of 11.22 Mbps and MAE of 6.45 Mbps for the considered dataset in the literature.

For further research directions, the distributed multi-agent learning approach can be extended to more complex prediction problem scenarios, e.g., grid-format and graph-format prediction problems with more varied neural network structures in a distributed way [47]. Another research direction which can be further explored is that the time-varying or learnable weight matrix in the interaction network can be further explored, instead of using the empirical predefined values.

Author Contributions: Conceptualization, W.J. and M.H.; methodology, W.J. and M.H.; software, W.J. and M.H.; validation, W.J. and M.H.; formal analysis, W.J. and M.H.; investigation, W.J. and M.H.; resources, W.J. and M.H.; data curation, W.J. and M.H.; writing—original draft preparation, W.J., M.H. and W.G.; writing—review and editing, W.J., M.H. and W.G.; visualization, W.J., M.H. and W.G.; supervision, W.J. and W.G.; project administration, W.J. and W.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Fundamental Research Funds for the Central Universities.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is available here: <https://doi.org/10.21227/jw40-y336> (accessed on 24 November 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pugliese, R.; Regondi, S.; Marini, R. Machine learning-based approach: Global trends, research directions, and regulatory standpoints. *Data Sci. Manag.* **2021**, *4*, 19–29. [CrossRef]
2. Zhang, Y.; Ren, G.; Liu, X.; Gao, G.; Zhu, M. Ensemble learning-based modeling and short-term forecasting algorithm for time series with small sample. *Eng. Rep.* **2022**, *4*, e12486. [CrossRef]

3. Zhao, E.; Sun, S.; Wang, S. New developments in wind energy forecasting with artificial intelligence and big data: A scientometric insight. *Data Sci. Manag.* **2022**, *5*, 84–95. [CrossRef]
4. He, M.; Gu, W.; Kong, Y.; Zhang, L.; Spanos, C.J.; Mosalam, K.M. Causalbg: Causal recurrent neural network for the blood glucose inference with IoT platform. *IEEE Internet Things J.* **2019**, *7*, 598–610. [CrossRef]
5. Shankarnarayan, V.K.; Ramakrishna, H. Comparative study of three stochastic future weather forecast approaches: A case study. *Data Sci. Manag.* **2021**, *3*, 3–12. [CrossRef]
6. Ramakrishnan, N.; Soni, T. Network traffic prediction using recurrent neural networks. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 187–193.
7. Azzouni, A.; Pujolle, G. NeuTM: A neural network-based framework for traffic matrix prediction in SDN. In Proceedings of the NOMS 2018–2018 IEEE/IFIP Network Operations and Management Symposium, Taipei, Taiwan, 23–27 April 2018; pp. 1–5.
8. Jiang, W.; Zhang, L. Geospatial data to images: A deep-learning framework for traffic forecasting. *Tsinghua Sci. Technol.* **2018**, *24*, 52–64. [CrossRef]
9. Jiang, W. Internet traffic matrix prediction with convolutional LSTM neural network. *Internet Technol. Lett.* **2022**, *5*, e322. [CrossRef]
10. Duan, Y.; Chen, N.; Shen, S.; Zhang, P.; Qu, Y.; Yu, S. Fdsa-STG: Fully Dynamic Self-Attention Spatio-Temporal Graph Networks for Intelligent Traffic Flow Prediction. *IEEE Trans. Veh. Technol.* **2022**, *71*, 9250–9260. [CrossRef]
11. Jiang, W.; Luo, J. Graph neural network for traffic forecasting: A survey. *Expert Syst. Appl.* **2022**, *23*, 117921. [CrossRef]
12. Ke, S.; Liu, W. Distributed Multi-Agent Learning is More Effectively than Single-Agent. 2021. Available online: <https://europepmc.org/article/ppr/ppr419060> (accessed on 1 November 2022).
13. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
14. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
15. Jiang, W. Internet traffic prediction with deep neural networks. *Internet Technol. Lett.* **2022**, *5*, e314. [CrossRef]
16. Shi, J.; Leau, Y.B.; Li, K.; Park, Y.J.; Yan, Z. Optimization and decomposition methods in network traffic prediction model: A review and discussion. *IEEE Access* **2020**, *8*, 202858–202871. [CrossRef]
17. Lohrasbinasab, I.; Shahraki, A.; Taherkordi, A.; Delia Jurcut, A. From statistical-to machine learning-based network traffic prediction. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e4394. [CrossRef]
18. Jiang, W. Cellular traffic prediction with machine learning: A survey. *Expert Syst. Appl.* **2022**, *201*, 117163. [CrossRef]
19. Tran, Q.T.; Hao, L.; Trinh, Q.K. A comprehensive research on exponential smoothing methods in modeling and forecasting cellular traffic. *Concurr. Comput. Pract. Exp.* **2020**, *32*, e5602. [CrossRef]
20. Perveen, A.; Abozariba, R.; Patwary, M.; Aneiba, A. Dynamic traffic forecasting and fuzzy-based optimized admission control in federated 5G-open RAN networks. *Neural Comput. Appl.* **2021**. [CrossRef]
21. Boutaba, R.; Salahuddin, M.A.; Limam, N.; Ayoubi, S.; Shahriar, N.; Estrada-Solano, F.; Caicedo, O.M. A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities. *J. Internet Serv. Appl.* **2018**, *9*, 1–99. [CrossRef]
22. Bayati, A.; Nguyen, K.K.; Cheriet, M. Multiple-step-ahead traffic prediction in high-speed networks. *IEEE Commun. Lett.* **2018**, *22*, 2447–2450. [CrossRef]
23. Zhang, Q.; Mozaffari, M.; Saad, W.; Bennis, M.; Debbah, M. Machine learning for predictive on-demand deployment of UAVs for wireless communications. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab, 9–13 December 2018; pp. 1–6.
24. Sun, S.C.; Guo, W. Forecasting wireless demand with extreme values using feature embedding in gaussian processes. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; pp. 1–6.
25. Xia, H.; Wei, X.; Gao, Y.; Lv, H. Traffic prediction based on ensemble machine learning strategies with bagging and lightgbm. In Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 20–24 May 2019; pp. 1–6.
26. Aldhyani, T.H.; Joshi, M.R. Integration of time series models with soft clustering to enhance network traffic forecasting. In Proceedings of the 2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 23–25 September 2016; pp. 212–214.
27. Qiu, C.; Zhang, Y.; Feng, Z.; Zhang, P.; Cui, S. Spatio-temporal wireless traffic prediction with recurrent neural network. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 554–557. [CrossRef]
28. Abdellah, A.R.; Muthanna, A.; Essai, M.H.; Koucheryavy, A. Deep Learning for Predicting Traffic in V2X Networks. *Appl. Sci.* **2022**, *12*, 10030. [CrossRef]
29. Gao, Y.; Zhang, M.; Chen, J.; Han, J.; Li, D.; Qiu, R. Accurate load prediction algorithms assisted with machine learning for network traffic. In Proceedings of the 2021 International Wireless Communications and Mobile Computing (IWCMC), Harbin, China, 28 June–2 July 2021; pp. 1683–1688.
30. Chien, W.C.; Huang, Y.M. A lightweight model with spatial-temporal correlation for cellular traffic prediction in Internet of Things. *J. Supercomput.* **2021**, *77*, 10023–10039. [CrossRef]

31. Zhan, S.; Yu, L.; Wang, Z.; Du, Y.; Yu, Y.; Cao, Q.; Dang, S.; Khan, Z. Cell traffic prediction based on convolutional neural network for software-defined ultra-dense visible light communication networks. *Secur. Commun. Netw.* **2021**, 2021, 9223965. [[CrossRef](#)]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
33. Shen, W.; Zhang, H.; Guo, S.; Zhang, C. Time-wise attention aided convolutional neural network for data-driven cellular traffic prediction. *IEEE Wirel. Commun. Lett.* **2021**, 10, 1747–1751. [[CrossRef](#)]
34. Wang, Z.; Wong, V.W. Cellular Traffic Prediction Using Deep Convolutional Neural Network with Attention Mechanism. In Proceedings of the ICC 2022—IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; pp. 2339–2344.
35. Hu, Y.; Zhou, Y.; Song, J.; Xu, L.; Zhou, X. Citywide Mobile Traffic Forecasting Using Spatial-Temporal Downsampling Transformer Neural Networks. *IEEE Trans. Netw. Serv. Manag.* **2022**. [[CrossRef](#)]
36. Jiang, W. Graph-based deep learning for communication networks: A survey. *Comput. Commun.* **2022**, 185, 40–54. [[CrossRef](#)]
37. Wang, Z.; Hu, J.; Min, G.; Zhao, Z.; Chang, Z.; Wang, Z. Spatial-Temporal Cellular Traffic Prediction for 5 G and Beyond: A Graph Neural Networks-Based Approach. *IEEE Trans. Ind. Inform.* **2022**. [[CrossRef](#)]
38. Fang, Y.; Ergüt, S.; Patras, P. SDGNet: A Handover-Aware Spatiotemporal Graph Neural Network for Mobile Traffic Forecasting. *IEEE Commun. Lett.* **2022**, 26, 582–586. [[CrossRef](#)]
39. Abdullah, M.; He, J.; Wang, K. Weather-Aware Fiber-Wireless Traffic Prediction Using Graph Convolutional Networks. *IEEE Access* **2022**, 10, 95908–95918. [[CrossRef](#)]
40. Zhou, X.; Zhang, Y.; Li, Z.; Wang, X.; Zhao, J.; Zhang, Z. Large-scale cellular traffic prediction based on graph convolutional networks with transfer learning. *Neural Comput. Appl.* **2022**, 34, 5549–5559. [[CrossRef](#)]
41. Oliveira, D.H.; de Araujo, T.P.; Gomes, R.L. An Adaptive Forecasting Model for Slice Allocation in Softwarized Networks. *IEEE Trans. Netw. Serv. Manag.* **2021**, 18, 94–103. [[CrossRef](#)]
42. Zerveas, G.; Jayaraman, S.; Patel, D.; Bhamidipaty, A.; Eickhoff, C. A Transformer-based Framework for Multivariate Time Series Representation Learning. *arXiv* **2020**, arXiv:2010.02803.
43. Fawaz, H.I.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.A. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, 33, 917–963. [[CrossRef](#)]
44. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.
45. Wang, Z.; Yan, W.; Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1578–1585.
46. Fawaz, H.I.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.A.; Petitjean, F. Inceptiontime: Finding alexnet for time series classification. *Data Min. Knowl. Discov.* **2020**, 34, 1936–1962. [[CrossRef](#)]
47. Shao, Y.; Li, H.; Gu, X.; Yin, H.; Li, Y.; Miao, X.; Zhang, W.; Cui, B.; Chen, L. Distributed Graph Neural Network Training: A Survey. *arXiv* **2022**, arXiv:2211.00216.