

Approximation of regression-based fault minimization for network traffic

Chanintorn Jittawiriyankoon

Graduate School of Advanced Technology Management, Assumption University, Thailand

Article Info

Article history:

Received May 22, 2019

Revised Apr 1, 2020

Accepted Apr 13, 2020

Keywords:

Approximation

Deep learning

Error minimization

Network traffic

Regression-based prediction

ABSTRACT

This research associates three distinct approaches for computer network traffic prediction. They are the traditional stochastic gradient descent (SGD) using a few random samplings instead of the complete dataset for each iterative calculation, the gradient descent algorithm (GDA) which is a well-known optimization approach in deep learning, and the proposed method. The network traffic is computed from the traffic load (data and multimedia) of the computer network nodes via the Internet. It is apparent that the SGD is a modest iteration but can conclude suboptimal solutions. The GDA is a complicated one, can function more accurate than the SGD but difficult to manipulate parameters, such as the learning rate, the dataset granularity, and the loss function. Network traffic estimation helps improve performance and lower costs for various applications, such as an adaptive rate control, load balancing, the quality of service (QoS), fair bandwidth allocation, and anomaly detection. The proposed method confirms optimal values out of parameters using simulation to compute the minimum figure of specified loss function in each iteration.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Chanintorn Jittawiriyankoon,
Graduate School of Advanced Technology Management,
Assumption University,
88 Moo 8, Bang Na Trad Km 26, Bang Sao Thong District,
Samut Prakan Province, Thailand, 10570.
Email: pct2526@yahoo.com

1. INTRODUCTION

Currently, computer networks and network traffic requests have been promptly amassed in many types of software in the 5G data network. Thus, scientists have studied the impacts of this progression based on users' requests from the network. Nonetheless, the traffic based on data stream has been regarded as critical traffic from users. Alshafut and Thayanathan [1] inspect massive data traffic for empowering a spectrum sharing over wireless network. Several plans are mentioned to handle this network traffic. Authors also propose a traffic model to lower the delay for the user's requests. They take diverse implications associated to a gigantic size of traffic in practice. They consider the traffic flow, starting from the access point to the arrival of requests to the carriers. This research [2] compares four different techniques for network traffic prediction. They are multi-layer perceptron with and without resilient back propagation, recurrent neural network, and stacked auto-encoding. Authors focus on time series traffic over the Internet and discover that both recurrent neural network and multi-layer perceptron approaches are better than the rest. The importance of content from service providers develops the network traffic through data center resources. The publication [3] proposes the employment of convolutional neural networks to estimate short-term variations in the traffic amount through the data center. The virtual

machine is used to configure the data center environment. Authors point out that the convolutional neural network achieves the non-linear network traffic, giving improvements to the data with specific standard deviation and mean absolute value. The traffic analysis and prediction are the important meaning of secured and reliable computer network. Various techniques have been proposed for network congestion analysis such as regression, deep learning etc. but many works in the area of traffic analysis and forecast are presented in [4]. The paper investigates, summarizes and reviews multiple publications. Although time series, and real-time network traffic approximation is obliged in various types of network software such as traffic management and resources allocation, but the searching for a predictor with high accuracy but low power consumption has been outlined in [5]. Predictors include three different classes such as neural networks, time series, and wavelet transform-based predictors. The analysis is based on real network indicators. The computation cost, and power consumption are listed. They follow the exponential function in terms of tradeoff between overhead cost and performance. The work [6] centers on the design, the estimation, and the analysis of the behavior of training models for forecasting the throughput of a single channel. A regression and fuzzy models are adopted for the estimation. Based on real network experiment of different channels, the impact of parameters on the approximation error is discussed. Results demonstrate that training models deliver accurate throughput estimation.

The Internet traffic approximation with three models of deep belief network is described in [7]. The neural network with four layers depth in each model to evaluate the non-linear and time series of Internet traffic is developed. The deep learning approach is used with unsupervised pre-processing of these layers. The method achieves estimation accuracy as well as a low figure of root-mean-squared error on given datasets. Wang et al. [8] present a deep learning model-based Internet traffic prediction. The approach takes integral correlations and traffic flow data into account. De-noising and encoding model is employed to observe Internet traffic characteristics, and is trained by a greedy algorithm. The estimation model, which is a part of the traffic scheduling system helps increase the bandwidth utilization of Internet network. Currently, computer network confronts with a gigantic traffic demand to handle up to the standard quality to users. An accurate network development is crucial to sustain revenues by shortened profit from a bandwidth usage.

Svigelj et al. [9] demonstrate a user-oriented tactic to computer network and model a network traffic to optimize a new service. The proposed method based on the end users and their profiles which can be gathered from real environment. The proposed method confirms that during experimental period the load differs less than 5% from the real figures. Song et al. [10] classify how assorted noise influences the performance. They apply stochastic gradient descent (SGD) for given datasets with noise and determine that it depends on the learning rate value. They then propose a mechanism for altering the learning rate and conduct an experiment on real environment to display that the proposed method is healthier than applying a fixed learning rate. For any tools based on performance evaluation, the paper [11] suggests to make use of SNMP software with a specified polling time. However, sampling too frequent causes network units to be overloaded and escalate network traffic. Too extended sample intervals slip the helpful information in the performance metrics. They also suggest to quantify the performance regarding user's network scope.

Computer network is becoming users' requirement, regarding software and users account. To develop a high-quality service, network analyst observes multiple aspects of the network traffic like channel traffic amount. When network's size grows, the monitoring task becomes crucial. The problem of using network traffic on particular channel to estimate the traffic on others is discussed in [12]. Although the approach costs more, but this method can obtain essential information to denote the network structure, and can help raise the results of estimation. A novel technique based on a computing node reserving the communication history is presented in [13]. Authors evaluate recent records in each node while disregarding uneven history then applying weight to consistent records. The technique per se prevents computing nodes from being struck by incorrect and hidden information. The technique reduces an error of network prediction delay about 50% using the evaluation based on real experimental data. Another latency prediction is proposed by [14]. The approach focuses on static and dynamic latency prediction based on time series 3D matrices.

Authors introduce a decomposition algorithm to divide latency matrices into a sub-component of the distance and network, and then strengthen the pattern of 3D data to upturn prediction accuracy. Experiments based on real monitoring proof that the method outperforms traditional estimation approaches. In order to analyze mobile data traffic, the algorithm based on machine learning is implemented [15]. The data-driven experiment on the performance prediction points out that a Markov predictor outperforms the traditional techniques in many cases. The proposed method achieves the accuracy of 70%, which outperforms all existing ones and improves 1% to 5% by machine learning. The latest success of deep learning supports novel tools that challenge problems in mobile infrastructure. To seal the gap between deep learning and mobile network, a survey of the boundaries is presented in [16]. Authors familiarize the background and deep learning approaches with prospective software. They also deliberate various techniques that ease the effective placement of deep learning on mobile networks. An in-depth analysis of mobile networking research based on deep learning is also listed. To shape deep learning to mobile systems is discussed.

Finally, they determine current confrontation. Tools that enable end-users with real time analysis and a quality of experience recommendations are explained in [17]. Authors also describe an end-user-centric system that empowers users to compile opinion scores and network traffic figures about the performance of services. The system includes cross-correlation measurement and provides prediction models for various services such as end-user profiles, quality of experience, and statistics. In conclusion, authors emphasize the research results, future directions, and challenges. Internet traffic prediction is also presented in [18]. Collected data from measurement is executed. Period is specified every day at the same time to collect those data. The experiment depends on the number of copies of identical resource from multiple servers. Paper describes legacy method, and preliminary data analysis. Prediction model and the dialogue of results are outlined as well. A data mining approach to approximate incoming, and outgoing data rate in networking environment based on association rule is presented in [19]. The mentioned data rate and bandwidth are performance metrics which can help sort out network traffic problems.

Author shows that this approach can estimate network traffic and can calculate data congestion and loss. Moreover, the method is applicable for routing allocation and performance upgrade. Approximating the performance of network traffic during a specified time with a possibility empowers improved routing selection, and data transfer that is critical in big data environments. A performance approximation model based on time series method, to increase the efficiency of resources utilization, scheduling, and data management is proposed in [20]. Authors create adjustment processes for identifying the patterns, period, diagnosis, and adjustment. They also demonstrate superior performance estimation in the adjustment model comparable to previous time series models. The non-linear data analysis based on echo state network to forecast the data traffic is presented in [21]. The approach is a new method based on recurrent neural networks, where a reservoir is issued and a weight matrix is used. A better echo state network concerning weakness on the activity function and the selection of matrix is proposed. The network outlines the scope of matrix and substitutes the activity in the middle layer of wavelet function.

Results prove that the proposed network is valid compared to traditional one. To design a traffic-aware network is important to model the traffic prediction and identify the traffic. The entropy theory to identify the traffic and validate the performance estimation is presented in [22]. A blueprint for traffic-based software to provide the forecast data is discussed. However, the above-mentioned works have been carried out on general predictions, but seldom shed light on the regression-based prediction for computer network traffic. Not to mention, with the explosively increasing demand for Internet access, there are imperative needs to estimate network traffic. To this persistence, it is crucial to minimize percentage of error for prediction. The rest of the paper is structured as follows. Section 2 offers traditional approximation methods as well as the proposed technique. Section 3 focuses on simulation results to validate the proposed method and the analysis. The conclusion and future direction of the research are discussed in section 4.

2. ESTIMATION METHODS

In this section, regression-based predictions of traffic flow on the computer network are discussed. The traffic matrix, holding flow figures among the computing nodes in computer networks, is insightful in several aspects of network control, such as traffic load allocation, traffic stability [23], load balancing, route selection, traffic shaping, outlining, and capacity forecasting. To operate a computer network efficiently it is often significant to quantify various attributes of the transmitted traffic (traffic flow). In the long run, measurements as such contribute optimization and network capacity planning. For instance, allocation of network traffic regarding source-destination pairwise leads a configuration of the computer network topology to take pairwise communication onto the similar path. A study of the network traffic between an overloading server, and its clients, is beneficial to apportion designated clients to an incremental server in a balanced fashion (load balancing). In the short operation, measurement of traffic loads helps tolerate the instant detection of misbehaviors. For instance, a sender announcing an unusual number of broadcast traffic possibly malfunctions. An interruption of traffic on source-destination pairwise designates an inconsistent usage of computer resources, like interactive game or gigantic file transfer, which rather avoids peak periods. In order to examine traffic loads accurately it is to measure all packets. However, it directs to data storage complications and ingests processing power. For that software, it needs to inspect packets using statistical method to approximate the actual counts. Stochastic sampling is perhaps preferable for the purpose of appropriateness and is desirable on grounds of accuracy as a complete packet sampling is tedious, and yet leads the traffic at peak. Not to mention it may exceed the capacity of the resources to keep up with, then packet loss steadily occurs at times.

Gradient Descent (GD) is a common optimization approach and can be applied for all learning algorithms in deep learning. Statistically, it uses the partial derivation of the parameter set associated to its inputs. The more the gradient (slope) is, the higher the slope will be. GD is a bulging function and an iterative computation to determine the figures of the input parameter function that minimizes the loss function as far as possible. The parameters are primarily clarified as a specific value then GD has been executed in multiple repetitions to determine the optimal figure of the input parameters, based on fundamental calculus, to expect the lowest figure

of the specified loss function. In this paper, two types of GD algorithm are concerned, stochastic gradient descent (SGD) and gradient descent algorithm (GDA).

2.1. Stochastic gradient descent (SGD)

It is an algorithm that is associated with a stochastic (random) probability. In SGD algorithm [24], some samples are chosen randomly instead of the complete samples from dataset for each repetition. In the classical GD algorithm, the complete dataset is taken into account. Although, using the entire dataset is actually meaningful for reaching the minima in a noisy pattern (environment), but the problem ascends as dataset size grows. In case of huge dataset, if classical GD approach is applied then it must take all samples to complete one round of repetition. Apparently, the computation cost is on the rise. The chaos ends by employing SDG. In this regard, it only performs selective and shuffle samples for each repetition.

2.2. Gradient descent algorithm (GDA)

Optimizing variables is the objective of all deep learning algorithms. The optimum figure of the gradient (slope) and the intercept are computed to get the best approximation in linear regression function. The objective function formulates input parameters onto output values. This applies for all regression predictions. Deep learning algorithm has coefficients that depict the algorithm approximation for the objective function. Individual algorithm has different coefficient however, an optimization to determine the set of coefficients, that represent the best approximation is expected. Optimization based on GD can apply for an algorithm with the set of coefficients, such as logistic or linear regression problems. The estimation of how well a learning model fits the objective function can be computed through different steps. The loss function represents how fit the model is performing on the training dataset. If the loss is extraordinary, the approximation is diverging too far from the target data. Thus, in deep learning algorithm, the decisive goal is to minimize the loss linked with the learning model. The loss function includes weighing the coefficients in the learning model by computing an approximation for each training dataset, and then associating the estimations to the output variables to determine an average error. The loss is computed for the algorithm via the training dataset for each step of the GDA [25], which is a popular algorithm in deep learning method.

2.3. Proposed method

Current learning models include datasets with attributes holding various parameters. As such it is problematic to determine the convexity of the loss function for the investigation. A novel dispersed algorithm for GD in the interpolating bound is presented in [26]. The algorithm associates to a simple distributed loss function. However, the use of GD is driven by the computation cost of executing back propagation via the whole training sets. The cost leads to clumsy convergence. In this paper fast convergence is proposed through datasets standardization. Standardizing data characteristics about zero, and the midpoint with a standard deviation of one is used when measurements have dissimilar units. Parameters that are collected at diverse scales do not accord similarly to the investigation and lead to an unfairness.

The proposed algorithm revises the present figure O_y^t based on the moving direction Δ_t by step value α_t to the recursive solution O_y^{t+1} . Then the recursive calculation of slope can be shortened in (1) shown in the below iteration.

$$O_y^{t+1} = O_y^t - \alpha_t \nabla F(O_y^t) \quad (1)$$

The goal is to minimize error in estimation of the function F, a humble technique to select the step value in fashion that minimizes the figure of future direction, i.e. compute the value that will be able to minimize $F(O_y^{t+1})$. Thus, the step value is calculated by (2).

$$\alpha_{t+1} = \arg \min_{\alpha \geq 0} F(O_y^t - \alpha \nabla F(O_y^t)) \quad (2)$$

Let $F: C \rightarrow K$ denote a convex function with parameters q, Q and the step size for recursive analysis along direction d is computed by the third order expansion of Taylor function as shown (3).

$$F(O_y^t) - \frac{1}{2q} \nabla_{d,d}^2 F(O_y^t) - \frac{1}{6q} \nabla_{d,d,d}^3 F(O_y^t) \leq \alpha^* \leq F(O_y^t) - \frac{1}{2Q} \nabla_{d,d}^2 F(O_y^t) - \frac{1}{6Q} \nabla_{d,d,d}^3 F(O_y^t) \quad (3)$$

for a recursive step r and $\xi > 0$, given that $F(O_y^r) - \min_{O_y^r \in C \rightarrow K} F(O_y^t) \leq \xi$ then we have;

$$r \geq \log\left(\frac{F(O_y^0) - \alpha_r}{\xi}\right) / \log\left(\frac{1}{1 - \frac{q}{Q}}\right). \quad (4)$$

The proposed algorithm is simplified as shown in Figure 1.

```

Proposed Algorithm
Require: Input parameter matrix  $[D]_{xy}$  with  $x$  rows and  $y$  columns
Ensure:  $[D]_{xy}$ ,  $O_j^t$  = candidates in step  $t$ ,  $\alpha_t$  = step size,
set  $t \leftarrow 0$ , Initialize  $O_j^t$ ,

for  $I = 1$  to  $x$  do
for  $J = 1$  to  $y$  do
    Read  $[D]_{xy}$ 
    Standardize  $[D]_{xy}$ 
end for
end for

while  $\| \nabla F(O_j^t) \| \geq 0$  do

     $O_j^{t+1} = O_j^t - \alpha_t \nabla F(O_j^t)$ 
     $\alpha_{t+1} = \arg \min_{\alpha \geq 0} F(O_j^t - \alpha \nabla F(O_j^t))$ 
    /** A new best step size**/
    if  $\alpha_{t+1}$  is in the range specified by Eq.3 Then
         $t \leftarrow t+1$ 
    Else It is "divergent"  $\rightarrow$  Exit.
end while
return  $O_j^t$ 
    
```

Figure 1. Proposed algorithm

3. RESULTS AND ANALYSIS

For the experiment to validate how to minimize error for aforementioned algorithms, three datasets that represent computer network components and the structure are simulated. The objective is to find the regression-based estimation for network traffic data. Tables 1-10 depict simulation results. All datasets are imbalanced characteristic regarding classes and attributes function. We simulate the logistic regression to minimize the loss function. SGD is not a competitive algorithm for complex datasets. However, we observe that the association between the GDA and the proposed algorithm is similar to achieve the optimal values.

Table 1. The dataset#1 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
1.06	1.06	1.61
0.14	0.15	0.15
0.15	0.14	0.14
0.17	0.14	
0.19		
0.17		
0.2		
	Optimal Value	
(5.04, -4.06)	(4.69, -4.14)	(4.73, -4.17)

Table 2. The dataset#2 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
2.31	2.31	0.34
0.85	1.1	0.22
0.82	0.69	0.16
0.8	0.5	0.15
0.48	0.36	0.14
0.35	0.26	
0.37	0.2	
0.34	0.18	
0.17	0.17	
0.16	0.16	
0.15	0.15	
	Optimal Value	
(7.55, -6.17)	(7.37, -6.61)	(5.45, -4.79)

Table 3. The dataset#3 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
5.15	5.15	0.44
0.29	0.29	0.34
0.25	0.28	0.26
0.22	0.27	0.23
0.21	0.26	0.2
0.2	0.25	0.19
0.19	0.23	0.18
0.17	0.22	0.17
0.19	0.21	
0.18	0.19	
0.17	0.18	
	0.17	
	Optimal Value	
(N/A)	(4.01, 1.36)	(4, 1.37)

Table 4. The dataset#4 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
0.37	0.37	0.39
0.33	0.35	0.37
0.32	0.33	0.35
0.3	0.3	0.32
0.29	0.28	0.31
0.27	0.26	0.27
0.26	0.24	0.25
0.25	0.23	0.23
0.23	0.21	0.21
0.24	0.19	0.19
0.22	0.18	0.18
0.19		
0.2		
0.18		
	Optimal Value	
(N/A)	(4.87, 1.01)	(4.82, 1.01)

Table 5. The dataset#5 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
1.25	1.25	2.19
5.09	1.01	1.28
5.03	0.81	0.78
4.69	0.77	0.75
2.71	0.74	0.73
1.77		
2.01		
1.11		
1.62		
1.73		
0.87		
(N/A)	Optimal Value (N/A)	(N/A)

Table 6. The dataset#6 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
2	2	1.39
1.97	1.75	1
1.92	1.28	0.83
1.54	0.93	0.74
1.07	0.8	0.71
2.36	0.75	
1.9	0.72	
1.16	0.73	
0.75	0.75	
0.82	0.73	
1	0.8	
0.76		
0.82		
0.84		
(N/A)	Optimal Value (N/A)	(0.48, -0.2)

Table 7. The dataset#7 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
3.15	3.15	5.92
2.87	1.80	1.1
2.85	0.88	0.46
3.16	0.7	0.38
2.29	0.81	0.32
2.03	0.98	0.30
3.91	1.15	
2.2		
1.38		
1.12		
1.41		
(N/A)	Optimal Value (N/A)	(-2.3,0.4)

Table 8. The dataset#8 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
4.52	4.52	5.37
3.17	2.71	2.04
2.4	1.07	1.25
1.24	1.08	0.85
0.77	1.52	1.02
0.97	0.76	0.96
1.63	0.96	
0.8	1.02	
3.77		
1.65		
(N/A)	Optimal Value (N/A)	(N/A)

Table 9. The dataset#9 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
4.44	4.44	3.75
3.12	2.65	2.65
2.28	1.06	1.51
0.74	0.86	0.7
1.43		
2		
1.45		
2.3		
(N/A)	Optimal Value (-0.6, -1.2)	(0.25,0.05)

Table 10. The dataset#10 outcome sequence of loss function with step size = 10

SGD	GDA	Proposed
5.16	5.16	3.13
4.12	3.55	2
2.52	1.83	0.9
0.9	0.69	0.78
1.04		
0.72		
0.99		
1.79		
0.83		
(N/A)	Optimal Value (-0.02,0.09)	(-0.13, -0.96)

Not to mention, the proposed algorithm provides a fast convergence to optimization. As the simulations implicate dispersed data, the initialized step size is critical, and we cannot anticipate these algorithms to be modest for all datasets. The loss function trajectories and their optimal values (if positive) are tabulated. Results confirm that the proposed method aligns with the GDA for all datasets although it is inflexible to find the optimization in the experiments as listed in Tables 5 and 8. Nevertheless, we achieve final results many more than the GDA.

4. CONCLUSION

The proposed algorithm is comprehensively validated versus experiments for various dispersed datasets. The regression-based approximations are in worthy treatment with the simulation experiments,

resulting the validity and analogy compared to the GDA. Furthermore, comprehensive simulation results, including outcome sequences, clearly reveal the fast convergence of optimal values in many cases. An adaptive step size is in the focus of our future work.

REFERENCES

- [1] Alshafut A., Thayanathan V., "Traffic Predicting Model for Dynamic Spectrum Sharing Over 5G Networks," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 369-374, 2018.
- [2] Oliveira T. P., Barbar J. S., Soares A. S., "Computer network traffic prediction: a comparison between traditional and deep learning neural networks," *International Journal of Big Data Intelligence*, vol. 3, no. 1, pp. 28-37, 2016.
- [3] Mozo A., Ordozgoiti B., Gómez-Canaval S., "Forecasting short-term data center network traffic load with convolutional neural networks," *PLoS ONE*, vol. 13, no. 2, pp. 1-31, 2018.
- [4] Parveen H., Srivastava R., "Network Traffic Analysis and Prediction-A Literature Review," *International Journal of Research and Development in Applied Science and Engineering*, vol. 10, no. 1, pp. 1-4, 2016.
- [5] Iqbal M. F., Zahid M., Habib D., John L. K., "Efficient Prediction of Network Traffic for Real-Time Applications," *Journal of Computer Networks and Communications*, vol. 2019, pp. 1-11, 2019.
- [6] Zhani M. F., Elbiaze H., Kamoun F., "Analysis and Prediction of Real Network Traffic," *Journal of Networks*, vol. 4, no. 9, pp. 855-865, 2009.
- [7] Narejo S., Pasero E., "An Application of Internet Traffic Prediction with Deep Neural Network," *Conference: WIRN 2016, 26th Italian Workshop on Neural Networks*, 2016.
- [8] Wang W. et al., "A network traffic flow prediction with deep learning approach for large-scale metropolitan area network," *IEEE/IFIP Network Operations and Management Symposium*, 2018.
- [9] Svigelj A., Serbec R., Alic K., "Network traffic modeling for load prediction: a user-centric approach," *IEEE Network*, vol. 29, no. 4, pp. 88-96, 2015.
- [10] Song S., Chaudhuri K., Sarwate A. D., "Learning from Data with Heterogeneous Noise using SGD," *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, 2015.
- [11] Rady A. M., Zakaria H. M., "Computer Network Performance management using a Simple Network Management Protocol," *International Journal of Computing Academic Research*, vol. 6, no. 2, pp. 50-58, 2017.
- [12] Vaughan J., Stoev S., Michailidis G., "Network-Wide Statistical Modeling, Prediction, and Monitoring of Computer Traffic," *Journal Technometrics*, vol. 55, no. 1, pp. 79-93, 2013.
- [13] Ping F., McConnell C., Hwang J., "A Retrospective Approach for Accurate Network Latency Prediction," *Proceedings of 19th International Conference on Computer Communications and Networks*, 2010.
- [14] Liu B., Niu D., Li Z., Zhao H. V., "Network latency prediction for personal devices: Distance-feature decomposition from 3D sampling," *IEEE Conference on Computer Communications*, 2015.
- [15] Chen G., "Spatiotemporal Individual Mobile Data Traffic Prediction," [Technical Report] RT-0497, INRIA Saclay-Ile-de-France, 2018. [Online]. Available: <https://hal.inria.fr/hal-01675573v2/document>
- [16] Zhang C., Paul P., Hamed H., "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys and Tutorials*, pp. 1-67, 2018.
- [17] Katsarakis et al., "On User-Centric Tools for QoE-Based Recommendation and Real-Time Analysis of Large-Scale Markets," *IEEE Communication Magazine*, vol. 52, no. 9, pp. 37-43, 2014.
- [18] Kaminska-Chuchmala A., "Spatial internet traffic load forecasting with using estimation method," *Procedia Computer Science*, vol. 35, pp. 290-298, 2014.
- [19] Prangchumpol D., "A Network Traffic Prediction Algorithm Based on Data Mining Technique," *International Journal of Computer and Information Engineering*, vol. 7, no. 7, pp. 999-1002, 2013.
- [20] Hu K., Alex S., Demetris A., Constantine D., "Estimating and Forecasting Network Traffic Performance Based on Statistical Patterns Observed in SNMP Data," *Proceedings of the 9th international conference on machine learning and data mining in pattern recognition*, 2013.
- [21] Ye D. et al., "Improved Echo State Network (ESN) for the Prediction of Network Traffic," *Proceedings of the 11th International Conference on Mobile Multimedia Communications*, 2018.
- [22] Li R., Zhao Z., Zhou X., Palicot J., Zhang H., "The prediction analysis of cellular radio access network traffic: From entropy theory to networking practice," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 234-240, 2014.
- [23] Zhang B., Mao Y., Zhang M., Yu Y., Jiang G., Deng B., "The Stability of the Internet Traffic Features," *Proceedings of International Conference on Education, Management and Computer Science*, 2016.
- [24] Loshchilov I., Hutter F., "SGDR: Stochastic Gradient Descent with Warm Restarts," *Proceedings of the 5th International Conference on Learning Representations*, 2017.
- [25] Chen C. S. et al., "Using Neural Network with Gradient Descent Adaptive Learning Rate Training Algorithm for Geometric Dilution of Precision Approximation," *IOSR Journal of Electrical and Electronics Engineering*, vol. 13, no. 4, pp. 1-6, 2018.
- [26] Mitra P., "Fast Convergence for Stochastic and Distributed Gradient Descent in the Interpolation Limit," *The 26th European Signal Processing Conference*, 2018.