Deep or Statistical: An Empirical Study of Traffic Predictions on Multiple Time Scales

Yu Qiao, Chengxiang Li, Shuzheng Hao, Jun Wu, Liang Zhang

Huawei Technology Ltd., Nanjing Research Center, China

Introduction

Traffic prediction aims to forecast the future traffic level based on past observations, which is a crucial technic in the field of network communication. Traffic prediction has a wide range of application, such as network congestion control and bandwidth allocation, which is a crucial technic in the field of network communications. There are briefly three categories of traffic prediction methods, which extract temporal patterns and inter-dependencies from the sequential data. Statistical models is a linear combination of sequences and noise factors; machine learning models regard the time series analysis as a regression problem; deep learning is a black-box algorithm for extracting temporal patterns. In [5], Yi *et al.* studied the traffic forecasting on multiple time scales via statistical models. Newly developed tools such as LSTNet [3] motivate us to compare the performance of different models on multiple time scales.

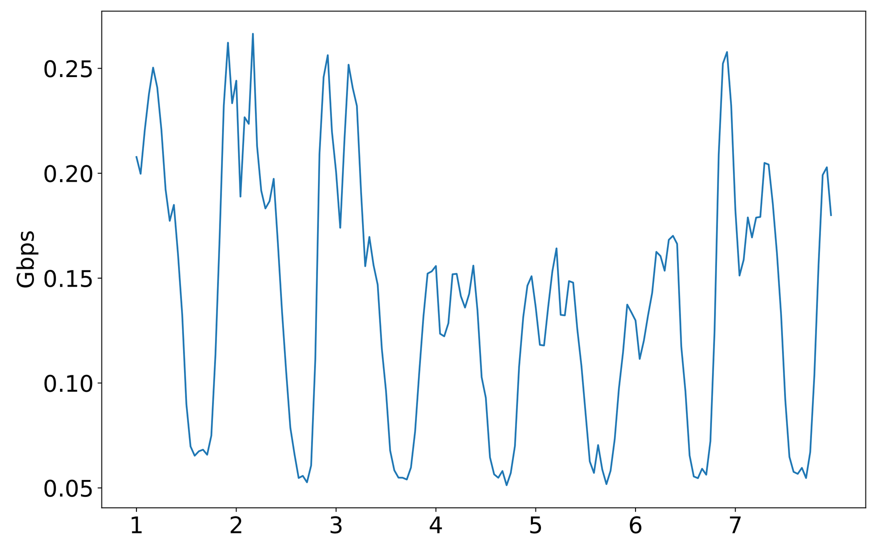
Trace

We consider a one-week real-world trace, which is captured on a 1-GigE campus link in the east of China. The trace is collected for one week from mid-September 2021. The link aggregates approximately 40 subscribers of student dormitories. We aggregate the trace for traffic intensity on six different time scales from 1 hour to 10 Seconds, shown in Figure 1. Due to the space limit, only two scales are presented. All traffic intensities are measured by Gbps.

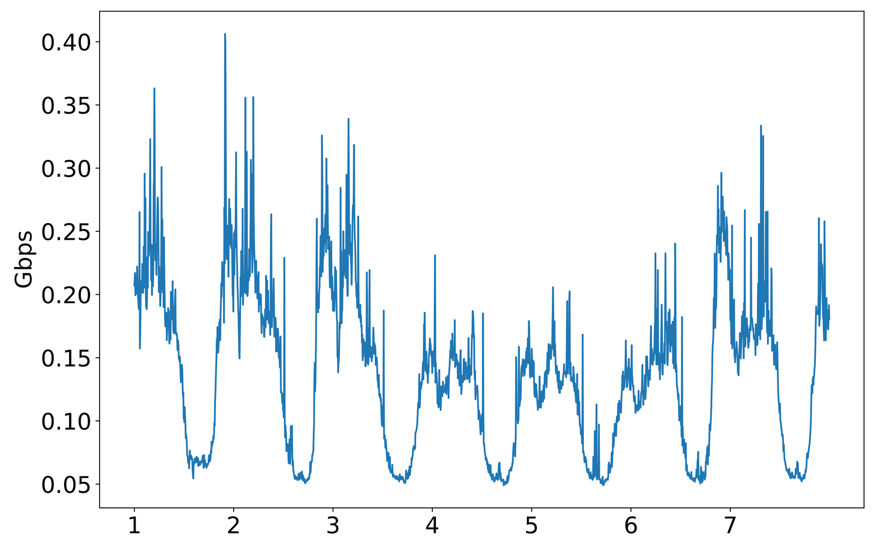
Table 1: Statistics(H:Hour,M:Minute,S:Second)

Scale	1H	30M	15M	5M	1M	10S
Length	168	336	672	2016	10080	60480
STD	0.060	0.061	0.062	0.063	0.066	0.073
SamEntro	0.79	0.67	0.46	0.37	0.43	0.56

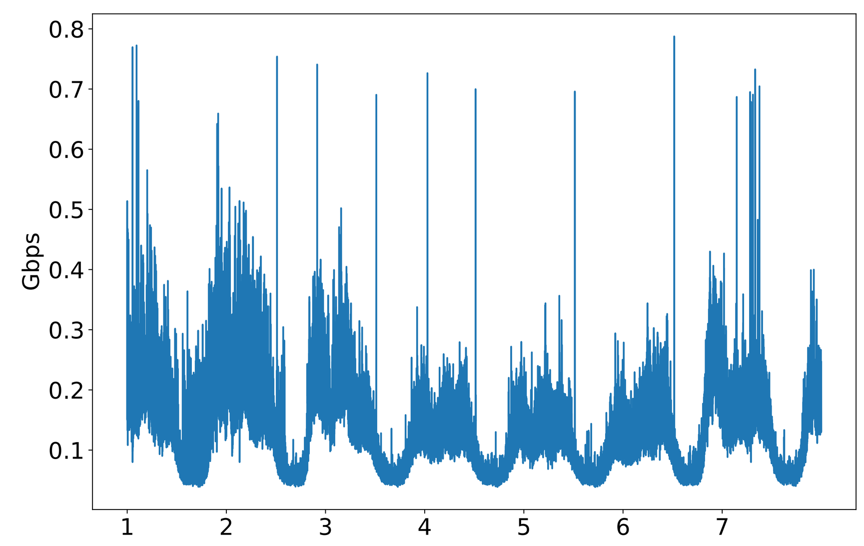
At different time scales, the traffic intensity presents different statistical characteristics, and the statistics are presented in Table 1. The concept of sample entropy was proposed in [4], which can measure the temporal predictability of a sequence. A smaller entropy means a stronger self-similarity and predictability. The granularity of 5 minutes shows a stronger self-similarity, and one hour's granularity has the largest entropy.



(a) One Hour



(b) Five Minutes



(c) Ten Seconds

Figure 1: Trace visualization of multiple time scales

Evaluation

	ARIMA	Ridge	XGB	RNN	LSTNet
1H	**0.1177**	0.1446	0.2192	0.1934	0.1848
30M	0.0914	0.0990	0.1002	0.0750	**0.0737**
15M	0.0886	0.0921	0.0861	0.0674	**0.0610**
5M	0.0796	0.0833	0.0833	0.0526	**0.0492**
1M	0.1019	**0.0815**	0.0840	0.1187	0.0994
10S	**0.1154**	0.1289	0.1213	0.1321	0.1188

Setups

We use 5 days' traffic for training and 2 days' traffic for testing. As for the prediction horizon, we use 20 samples for a one-step forecasting. Mean absolute percentage error (MAPE) is applied as the metric, formulated as:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i' - y_i|}{|y_i|} \tag{1}$$

where y_i' and y_i denote the predicted and true values, and n means the size of test sets. We select 7 representative time series forecasting algorithms to compare their performances on multi-time scales.

1. For a one-week trace, the granularity of 5 minutes, which has the smallest entropy, has the smallest MAPE via deep models compared with other granularities and algorithms, and it shows a stronger predictability.
2. Deep learning is ideal for a coarser granularity, and it verifies that deep models can learn complicated periodic features. However, the performances of deep learning such as RNN will decrease dramatically, when data is not enough.
3. Statistical models are ideal for a finer granularity, it can capture short-term sequential patterns. Also, statistical models such as ARIMA outperform when the training data is insufficient.

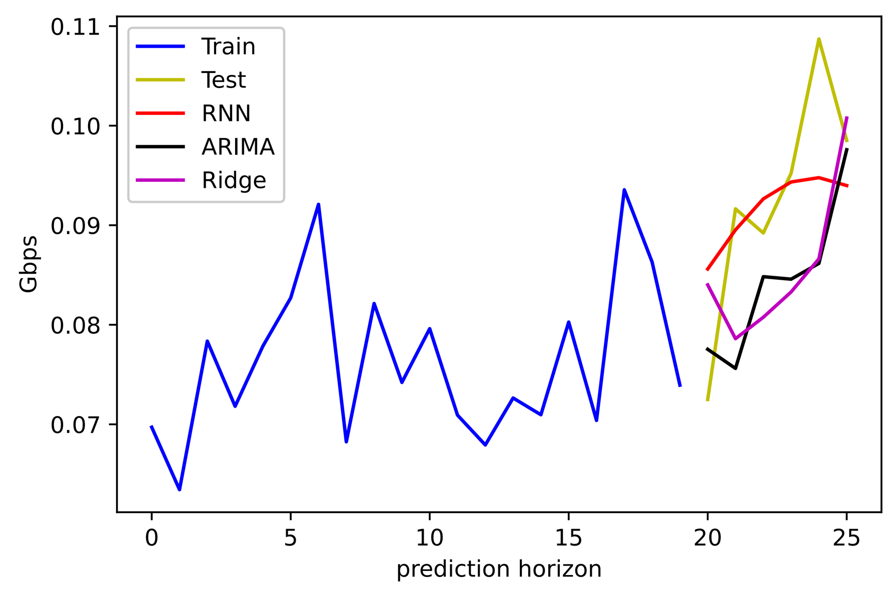


Figure 2: Outputs of different models

In Figure 2, we show what different models learned with the granularity of 5 minutes. The size of training samples are 20, and prediction horizon is 6. Predictions are made in a recursive manner. From Figure 2, we observe that RNN can capture the periodic patterns of the sequence, and it failed to get the short-term fluctuations, and the predictions seem to have been smoothed. In the contrast, the forecasting of regressive and statistical models can reflect some short-term changes, but it does not perform well for the trend compared to deep models.

References

[1] T. Chen and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In ACM.

[2] R. F. Engle. 1981. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. In Econometrica, Vol. 50. 987–1008.

[3] Lai Guokun, Chang Wei Cheng, Yang Yiming, and Liu Hanxiao. 2018. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. ACM (2018).

[4] Richman Joshua S. and J. Randall Moorman. 2000. Physiological time- series analysis using approximate entropy and sample entropy. Ameri- can Journal of Physiology-Heart and Circulatory Physiology (2000).

[5] Q. Yi, J. Skicewicz, and P. Dinda. 2004. An empirical study of the multi- scale predictability of network traffic. In IEEE International Symposium on High Performance Distributed Computing.