Comparison of Network Traffic Forecasting Schemes

Chaeran Kim, Chong-Kwon Kim*
Department of Energy Engineering
Korea Institute of Energy Technology(KENTECH)
Naju-si, Jeollanam-do 58217, Republic of Korea
cofks05, ckim@kentech.ac.kr

***Abstract*—The network traffic prediction or forecasting problem has attracted tremendous interest from both academia and industry because traffic prediction is widely used as the fundamental data in establishing or in updating network infrastructure. Recently, numerous clever time-series analysis schemes based on machine learning techniques such as RNNs (Recursive Neural Networks) and self-attention architecture have been developed. However, most of the previous studies in the area of artificial intelligence (AI) do not deal with network traffic in their experiments. In this paper, we deal with the network traffic prediction problem using the domain knowledge accumulated in the field of networking. In particular, we define channels for network traffic and interpret temporal and spatial information from the real-world traffic dataset. Our experiments on the Abilene dataset show that the recent patch mixing model performs best in terms of prediction error. In addition, we observe that both RNN models and simple transformer-based models show inadequate performances.**

***Index Terms*—Time-series forecasting, Transformer, Network traffic, Patch, Mixing**

I. Introduction

The network traffic has experienced explosive growth due to the prevailing penetration of video and multimedia services. Because network traffic increases rapidly and abruptly, ISPs (Internet Service Providers) are urged to act proactively to satisfy rapidly growing customer demands. Network traffic forecasting can be classified into two categories in terms of the length of forecasting: long-term prediction and short-term prediction. Long term predictions [1, 2] forecast traffic in remote future ranging from several months or several years later in the future. Long term predictions play an important role in network planning tasks such as network topology design and network capacity planning. In contrast, short-term network predictions [3, 4, 5] cover short temporal spans such as millisecond to minute scales. Short term predictions are used for congestion control, routing, network resource management and various dynamic network control tasks.

Because network traffic prediction is essential for both dynamic network control and for long term network planning, the literature of network traffic prediction is prominent in both breadth and depth [6, 7, 8, 9]. The traffic forecasting methods can be classified into four classes in terms of foundation techniques: linear and non-lines time- series forecasting, conventional machine learning, deep learning, and transformer based methods. Even though conventional linear/non-linear models and data driven models provision satisfactory performances, recent schemes based on deep learning and self-attention architecture generally provide better performances. Particularly, recent transformer based models are shown to be superior to non-transformer models in long term forecasting. Therefore, in this paper, we focus on deep learning based models and transformer models.

We select four representative forecasting techniques and compare their performance in predicting future network traffic. The four methods are LSTM (Long Short Term Memory), Informer, Autoformer, and TSMixer. Informer and Autoformer include self-attention modules in their networks. Furthermore, the TSMixer model adopts patching and mixing techniques developed in the area of computer vision. We conducted a performance analysis using the data set obtained from the real-world network called the Abilene network. The Abilene network is a high-performance backbone network created by the Internet2 community. We pre-process the raw Abilene dataset to make them suitable for time-series forecasting tasks.

Our extensive experiments on the Abilene dataset show that TSMixer provides significantly superior performances over the other three methods. However, somehow contrary to prior studies, we observe that transformer architecture models, Informer and Autoformer, show poor performances; their performances are worse than that of LSTM. This may indicate that for network traffic prediction, we may need to devise additional novel components in addition to transformers.

Our contributions can be summarized as follows.

- We apply the most recent time-series forecasting methods such as TSMixer for network traffic forecasting. As far as we know, this is the first application of the techniques to networking domain.
- We perform extensive experiments on real-world dataset, Abilene traffic data. We preprocess the traffic data in order to make them suitable for time-series forecasting.
- We discover that methods which combine patching and mixing techniques are superior to other models.

The rest of the paper is organized as follows. Related work section explains prior methods. We classify prior time-series analysis models into four categories. Time-Series Forecasting Models section describes the four techniques analyzed in this paper. Especially we explain the TSMixer model in a greater detail because it is the best performing method. Experimental Results section illustrates the details of Abilene traffic dataset and performance evaluation process. In addition, the results obtained from our experiments are described in this section.

II. Related Work

This section describes four classes of forecasting techniques: linear and non-linear time-series forecasting, conventional machine learning, deep learning, and transformer based methods. Early studies utilized various linear and non-linear forecasting techniques including ARMA (Auto-Regressive Moving Average) [10], ARIMA (Auto-Regressive Integrated Moving Average) and SARIMA (Seasonal ARIMA)[12, 13]. As big data on real-world traffic including GEANT [14] and Abilene networks [15, 16] have been accumulated and are open to the public access, many big-data based network traffic prediction schemes have been proposed. The big data based models include conventional machine learning techniques; linear and non-linear regression, SVM (Support Vector Machine) [17], Random forest, HMM (Hidden Markov Models) [18, 19], and Wavelet transform [20].

Rapid advances in NNs (Neural Networks) and DL (Deep Learning) have changed the research landscape of time-series forecasting and prediction significantly. Starting from the basic CNN (Convolutional NN) and RNN (Recurrent NN) [21], their variations such as LSTM (Long and Short Term Memory) [22, 23] and GRU (Gated Recurrent Unit) [24, 25] have been widely and successfully applied to traffic forecasting. Even though RNN based models provide more accurate prediction than traditional machine learning techniques, their performances are worse than the conventional methods in certain cases, indicating their lack of generality.

More recently, as the self-attention mechanism [26] have became popular and widely used in NLP and computer vision areas, several methods founded over the transformer architecture have been proposed. These include Informer [27], Autoformer [28] and Fedformer[29]. Informer is one of early method that utilized the transformer architecture. Informer proposes a module so called ProbSparse self-attention mechanism aimed to reduce the time complexity and memory requirements of transformers. In addition, Informer adopts self-attention distillation and generative decoder to enhance the accuracy of forecasting. Autoformer overcomes the drawback of sparse point-wise self-attention by introducing a decomposition architecture with an Auto-Correlation mechanism. Autoformer combines Transformer with seasonal trend decomposition to capture the global view of time-series. Transformer based methods achieve better performances than non-Transformer methods in long-term forecasting, but several recent studies [30, 31, 32] reported that these methods perform worse than traditional RNN methods in certain environments.

III. Time-Series Forecasting Models

This section explains the four time-series forecasting methods that we selected for performance analysis: LSTM, Informer, Autoformer, and TSMixer. LSTM is a variant of RNN designed to capture long-range dependencies in time-series data. LSTM also relieves the vanishing gradient problem. Unlike traditional RNNs, LSTMs use a unique architecture that includes memory cells and gating mechanisms. As shown in Fig. 1, the input, forget, and output gates regulate the flow of information such that important contexts are propagated over long future sequences. LSTM has enjoyed successful results in the area of NLP (Natural Language Processing) where sentences are considered as series of words. This ability makes LSTMs ideal for tasks like speech recognition, language modeling, and time-series prediction.

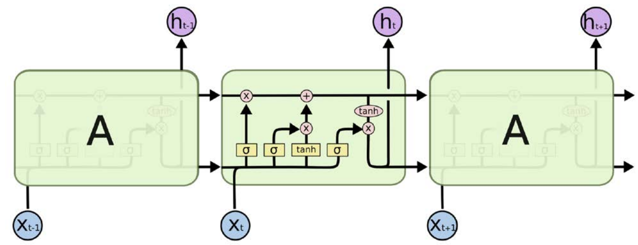
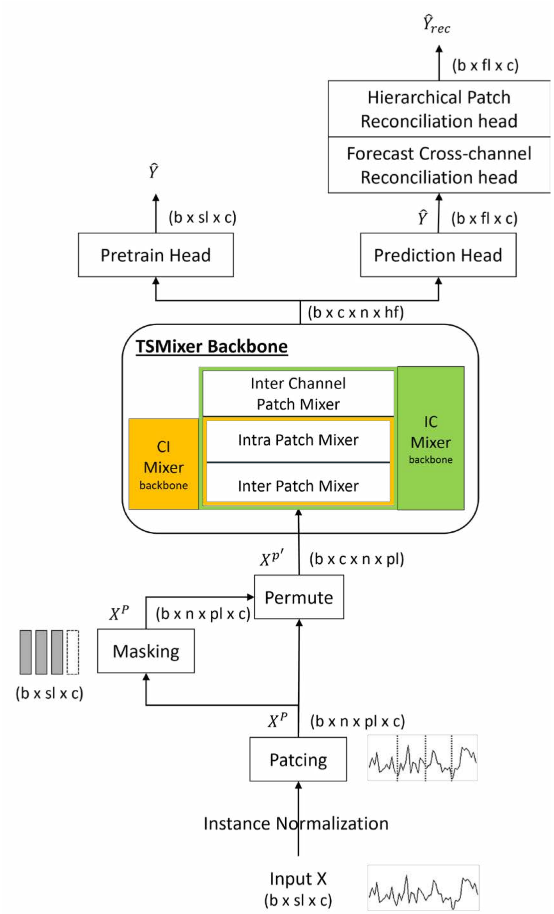


Fig. 1: Architecture of LSTM

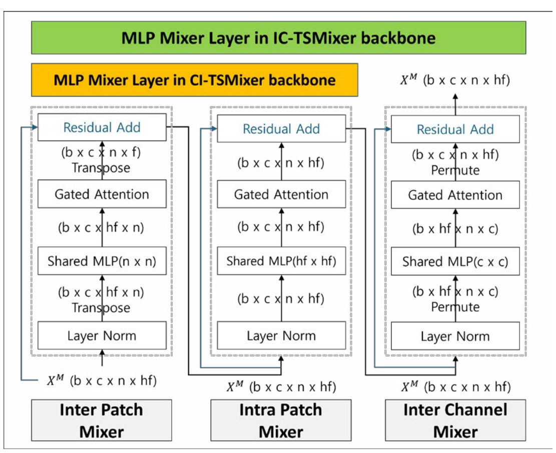
Informer and Autoformer, built on the transformer architecture, are designed to solve limitations of traditional methods like RNNs in long-term time-series forecasting. For effective capturing of temporal patterns in long time sequences, they use optimized attention mechanisms. Even though the methods based on transformer architecture provide good performances, they incurs large computational overhead. In result, plain transformer scheme suffers from the scalability problems because model sizes increase quadratically with the sequence length. Informer adopts two unique mechanisms, a probabilistic attention mechanism and a distilled attention, that select informative temporal intervals. Ignoring less informative intervals, Informer reduces the computational complexity and enhances the scalability of the model significantly.

Like the Informer, Autoformer is another Transformer-based model specifically designed for time-series data with seasonal patterns. To segregate trend component and seasonal component in time-series data, Autoformer uses a decomposition block. The separated components are then further processes by a conventional transformer and seasonal attention mechanism, respectively. This decomposition allows Autoformer to better capture both short-term (seasonal) and long-term (trend) dependencies separately, improving prediction accuracy for complex time-series forecasting tasks. Prior experimental studies show that both Informer and Autoformer significantly improve forecasting accuracy, scalability, and efficiency over RNN-based methods.

TSMixer is a recent methodology that applies ViTs (Vision Transformers) developed in the computer vision area to time-series forecasting tasks. Like the ViT schemes, they use the concept of ”patches” to capture spatial-temporal patterns in time-series data. While a patch in the computer vision context



(a) Overall Architecture



(b) Details of backbone architecture

Fig. 2: Architecture of TSMixer

is a sub-region in an image, a patch in time-series applications is a small temporal interval. Patching has a potential to reduce computational overheads significantly. Unlike the traditional transformer architecture that deals with long time-series, PatchTST partitions the original time-series into patches and applies mixing technique. This approach can reduce the computational overheads and enhance the scalability of the model significantly. In addition to the concept of patching, TSMixer (Patch Time-Series Mixer) adopts feature across both temporal and spatial dimensions. The architecture of TSMixer is illustrated in Fig. 2.

IV. Experimental Results

We perform extensive experiments with four time-series forecasting schemes on real-world network traffic dataset. The dataset is the Abilene dataset obtained from the real Abilene network. The Abilene network consists of 12 backbone nodes. The raw Abilene data are a traffic matrix (TM) of 12 by 12 size. Each component of the TM is the traffic load from source (row) to destination (column) measured at every five minute. We need to introduce the concept of channel to fully enjoy the features of transformer. We compute ingress traffic at each destination node by adding the source-destination components from 12 source nodes. After the pre-processing, we obtain 12 ingress traffic loads and use them as channels.

Throughout the paper, we use the following variable names:

- $\mathbf{X}_{sl\times c}$: A multivariate time-series of length sl and c channels.
- sl: Length of the input sequence.
- fl: Length of the forecast sequence.
- b: Batch size.
- n: Number of patches.
- pl: Patch length.
- hf: Dimension of hidden features.
- nl: Number of MLP-Mixer layers.
- $\mathcal{M}$: Learned DNN model.
- $\hat{Y}_{rec}$: Actual base prediction.
- Y_{rec}: Base ground truth.

The multivariate forecasting task involves predicting future time-series values based on a given historical data sequence :

$$\hat{\mathbf{Y}}_{fl\times c} = \mathcal{M}(\mathbf{X}_{sl\times c}). \quad (1)$$

The ground truth future values will be denoted by $\mathbf{Y}_{fl\times c}$. We use MSE (Mean Square Error) and MAE (Mean Absolute Error) as the performance metrics.

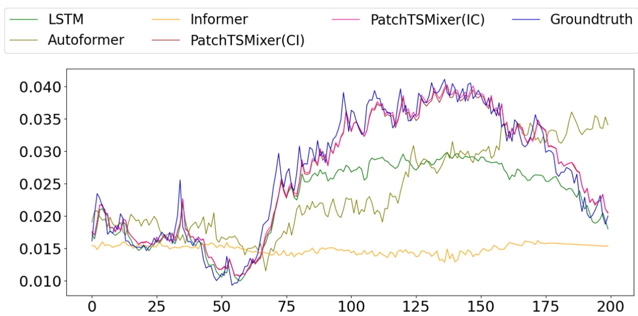
The performance of four time-series forecasting schemes is summarized in Table 1. We analyzed the performances varying the forecasting horizons; 1, 2, 4, 8 and 16. For example, a forecasting horizon of 4 means that we predicts traffic loads at 20 minutes (4 times 5 minutes) in the future. Note that the raw traffic and in our experiment, we use five minutes as a unit. As the horizon increases, the accuracy deteriorates.

In Table 1, we can observe that TSMixer performs best in all experimental environments. Surprisingly, LSTM performs better than Informer and Autoformer. This result may indicates that the transformer architecture may fail to perform adequately in some network traffic prediction tasks. We also can observe that accuracy of prediction deteriorates as the horizon length increases.

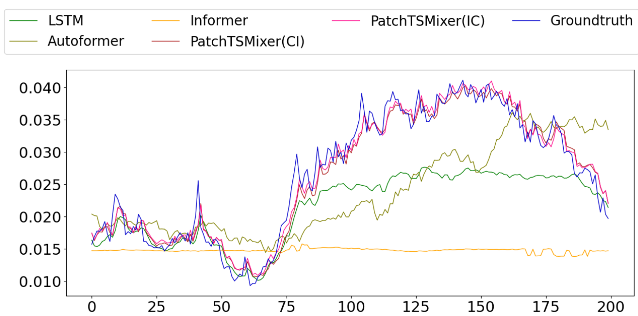
Figure 3 compares the actual traffic loads and predicted traffic loads at node 7. At both horizons of 1 and 8, TSMixer predicts traffic loads most precisely.

TABLE I: Performance comparison of different models on the Abilene network. The top results are highlighted in bold, and the second-best results are indicated with underlining.

Models		TSMixer(CI)		TSMixer(IC)		Autoformer		Informer		LSTM	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
horizon	1	**0.1237**	0.2159	0.1238	**0.2158**	0.6880	0.5345	1.0274	0.5873	0.2388	0.2922
	2	**0.1548**	0.2410	**0.1548**	**0.2405**	0.8176	0.5842	1.0420	0.5838	0.2961	0.3264
	4	0.2021	0.2722	**0.2009**	**0.2701**	0.8275	0.5871	1.1750	0.6348	0.3614	0.3538
	8	0.2749	0.3138	**0.2703**	**0.3091**	0.8488	0.5883	1.2426	0.6598	0.4411	0.3896
	16	0.3942	0.3733	**0.3791**	**0.3612**	0.8328	0.5713	1.4324	0.7162	0.5606	0.4255



(a) Horizon 1



(b) Horizon 8

Fig. 3: Actual and forecasted network traffic.

V. Conclusion

In this paper, we address the network traffic prediction problem. The time-series analysis has achieved significant improvements in recent years. Many clever algorithms utilizing the recent self-attention architecture, patching and mixing schemes have been proposed. We selected four representative time-series forecasting methods, LSTM, Informer, Autoformer, and TSMixer, and conducted extensive performance experiments on real-world data obtained from the Abilene network. Our performance study shows that TSMixer performs best among four schemes. Transformer-based models, Informer and Autoformer, fail to provide adequate performances. They perform worse than a traditional LSTM model indicating the necessity of further investigation to discover and solve the problem.

Acknowledgment

This work was supported by the KENTECH Research Grant(202200019A), by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(IITP-2024-RS-2022-00156287), and by Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(No.RS-2021-II212068, Artificial Intelligence Innovation Hub)

References

[1] Papagiannaki, Konstantina, et al. "Long-term forecasting of Internet backbone traffic." IEEE transactions on neural networks 16.5 (2005): 1110-1124.
[2] Zhang, Chaoyun, and Paul Patras. "Long-term mobile traffic forecasting using deep spatio-temporal neural networks." Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing. 2018.
[3] Vlahogianni, Eleni I., John C. Golias, and Matthew G. Karlaftis. "Short-term traffic forecasting: Overview of objectives and methods." Transport reviews 24.5 (2004): 533-557.
[4] Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. "Short-term traffic forecasting: Where we are and where we're going." Transportation Research Part C: Emerging Technologies 43 (2014): 3-19.
[5] Shao, Qitan, et al. "An adaptive composite time series forecasting model for short-term traffic flow." Journal of Big Data 11.1 (2024): 102.
[6] Joshi, Manish, and Theyazn Hassn Hadi. "A review of network traffic analysis and prediction techniques." arXiv preprint arXiv:1507.05722 (2015).
[7] Ferreira, Gabriel O., et al. "Forecasting network traffic: A survey and tutorial with open-source comparative evaluation." IEEE Access 11 (2023): 6018-6044.
[8] Lohrasbinasab, Iraj, et al. "From statistical-to machine learning-based network traffic prediction." Transactions on Emerging Telecommunications Technologies 33.4 (2022): e4394.
[9] Abbasi, Mahmoud, Amin Shahraki, and Amir Taherkordi. "Deep learning for network traffic monitoring and analysis (NTMA): A survey." Computer Communications 170 (2021): 19-41.
[10] P. K. Hoong, "BitTorrent network traffic forecasting with ARMA", Int. J. Comput. Netw. Commun., vol. 4, no. 4, pp. 143-156, Jul. 2012.
[11] I. K. T. Tan, P. K. Hoong and C. Y. Keong, "Towards forecasting low network traffic for software patch down-

loads: An ARMA model forecast using CRONOS", Proc. 2nd Int. Conf. Comput. Netw. Technol., pp. 88-92, 2010.
[12] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, J. Song, et al., "Big data driven mobile traffic understanding and forecasting: A time series approach", IEEE Trans. Services Comput., vol. 9, no. 5, pp. 796-805, Sep./Oct. 2016.
[13] M. Zhang, H. Fu, Y. Li and S. Chen, "Understanding urban dynamics from massive mobile traffic data", IEEE Trans. Big Data, vol. 5, no. 2, pp. 266-278, Jun. 2019.
[14] Koster, Arie MCA, and Manuel Kutschka. "Network design under demand uncertainties: A case study on the abilene and GEANT network data." Photonic Networks, 12. ITG Symposium. VDE, 2011.
[15] Almes, Guy, et al. "Abilene Comes Online: Building an Advanced IP Network." IEEE Internet Computing 2.6 (1998): 104.
[16] Zhang, Yin, et al. "Network anomography." Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement. 2005.
[17] N. Stepanov, D. Alekseeva, A. Ometov and E. S. Lohan, "Applying machine learning to LTE traffic prediction: Comparison of bagging random forest and SVM", Proc. 12th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT), pp. 119-123, Oct. 2020.
[18] H. Lu and F. Yang, "A network traffic prediction model based on wavelet transformation and LSTM network", Proc. IEEE 9th Int. Conf. Softw. Eng. Service Sci. (ICSESS), pp. 1-4, Nov. 2018.
[19] Z. Tian, "Network traffic prediction method based on wavelet transform and multiple models fusion", Int. J. Commun. Syst., vol. 33, no. 11, pp. e4415, Jul. 2020.
[20] T. Chadza, K. G. Kyriakopoulos and S. Lambotharan, "Contemporary sequential network attacks prediction using hidden Markov model", Proc. 17th Int. Conf. Privacy Secur. Trust (PST), pp. 1-3, Aug. 2019.
[21] Ramakrishnan, Nipun, and Tarun Soni. "Network traffic prediction using recurrent neural networks." 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018.
[22] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, et al., "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach", Proc. IEEE Conf. Comput. Commun., pp. 1-9, May 2017.
[23] X. Luo, D. Li, Y. Yang and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM", J. Adv. Transp., vol. 2019, pp. 1-10, Feb. 2019.
[24] S. A. Patil, L. A. Raj and B. K. Singh, "Prediction of IoT traffic using the gated recurrent unit neural network-(GRU-NN-) based predictive model", Secur. Commun. Netw., vol. 2021, pp. 1-7, Oct. 2021.
[25] R. Fu, Z. Zhang and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction", Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC), pp. 324-328, Nov. 2016.
[26] Vaswani, A. "Attention is all you need." Advances in Neural Information Processing Systems (2017).
[27] ZHOU, Haoyi, et al. "Informer: Beyond efficient transformer for long sequence time-series forecasting." Proceedings of the AAAI conference on artificial intelligence. p. 11106-11115. 2021.
[28] Wu, Haixu, et al. "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting." Advances in neural information processing systems 34 (2021): 22419-22430.
[29] Zhou, Tian, et al. "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting." International conference on machine learning. PMLR, 2022.
[30] Zeng, Ailing, et al. "Are transformers effective for time series forecasting?." Proceedings of the AAAI conference on artificial intelligence. Vol. 37. No. 9. (2023).
[31] Lin, Shengsheng, et al. "Segrnn: Segment recurrent neural network for long-term time series forecasting." arXiv preprint arXiv:2308.11200 (2023).
[32] Ekambaram, Vijay, et al. "Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting." Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023.