

Modelagem e Previsão de Tráfego de Rede: Uma Abordagem para ISP's

Murilo A. Chianfa¹, Bruno B. Zarpelão¹, Sylvio B. Junior²

¹Departamento de Computação – Universidade Estadual de Londrina (UEL)

²Departamento de Engenharia Eletrônica e Arquitetura – Universidade de Trieste (UNITS)

{murilo.chianfa,brunozarpelao}@uel.br, sylvio.barbonjunior@units.it

Abstract. *This study develops a method for network administrators to conduct extensive tests on traffic data using various time series forecasting models. Three algorithms were selected: SARIMAX, Prophet, and LSTM. The portal assists in selecting the best model, optimizing hyperparameters, and evaluating the trade-off between time and prediction efficiency.*

Resumo. *Este estudo desenvolve um método para que administradores de rede possam realizar testes extensivos em dados de tráfego utilizando diversos modelos de previsão de séries temporais. Três algoritmos foram selecionados: SARIMAX, Prophet e LSTM. O portal auxilia na escolha do melhor modelo, na otimização de hiperparâmetros e na avaliação do tempo versus eficiência das previsões.*

1. Introdução

Atualmente, com o advento de novas tecnologias no setor de telecomunicações, como Internet das Coisas (IoT), WiFi-7 e XGS-PON, estamos enfrentando um grande aumento no volume de tráfego de rede produzido como nunca antes visto [Fleischer, 2017]. Provedores de Serviço de Internet (ISPs) tentam implementar técnicas de monitoramento e análise de dados para tentar otimizar o fluxo de trabalho dessas redes, mas técnicas tradicionais têm se mostrado ineficazes para este trabalho.

Previsão de tráfego de rede desempenha um papel fundamental no avanço das redes de computadores, áreas como gerenciamento, design, alocação de recursos, engenharia de tráfego e detecção de anomalias [Troia et al., 2017]. Duas principais categorias de predição são atualmente empregadas, *long-term* e *short-term*. *Long-term* atualmente é mais utilizada para planejamento futuro de capacidade e recursos necessários, já *Short-term* sendo empregada para melhorar a qualidade dos serviços como Qualidade de Serviço (QoS) e alocação dinâmica de recursos [Andreoletti et al., 2019].

1.1. Objetivos da pesquisa

O principal foco desta pesquisa é desenvolver um portal para que administradores de rede possam realizar extensivos testes em seus próprios dados, utilizando diferentes modelos de previsão de séries temporais com diferentes parâmetros para cada modelo. Três principais perguntas poderão ser respondidas.

- Qual algoritmo estatístico ou de inteligência artificial melhor se adequa ao cenário do tráfego da minha rede e de meus clientes?

- Quais são as melhores combinações de hiper-parâmetros para meu tipo de tráfego?
- O tempo para gerar o modelo versus sua eficiência de predição são adequados ao meu caso de uso levando em consideração diferentes vetores da rede?

2. Trabalhos relacionados

Extensas pesquisas mostram que modelos estatísticos como ARIMA e Holt-Winters podem chegar a um bom resultado quando temos bastante dados do tráfego de rede e estes possuem uma boa sazonalidade. Muitos destes trabalhos comparam estes modelos com modelos de aprendizado de máquina, obtendo um melhor resultado quanto geral nos testes propostos pelos mesmos [Krishnaswamy et al., 2020; Madan and Mangipudi, 2018].

Alguns trabalhos utilizam-se especificamente de características temporais do tráfego de rede para categorizar e classificar ataques de rede DDoS [Halladay et al., 2022], outros utilizam uma abordagem híbrida para o mesmo propósito, também utilizando características temporais, colacionando-as com a entropia dos recursos dos fluxos IP observados [Ding, Savi, and Siracusa, 2022].

3. Escolha dos algoritmos

Para o trabalho proposto, foram selecionados três algoritmos, dois deles sendo modelos estatísticos SARIMAX, Prophet como um modelo de regressão ativa e, por fim, um modelo de rede neural autoregressivo LSTM. Todos estes se propõem a fazer previsões de séries temporais representada pela sequência de observações (y_1, y_2, \dots, y_t) . A previsão para um horizonte h períodos à frente pode ser expressa da seguinte forma [Adhikari and Agrawal, 2013].

$$(y_1, y_2, \dots, y_t) \xrightarrow{\text{Previsão}} \hat{y}_{t+h} = f(y_t, y_{t-1}, \dots, y_{t-p+1}) + \epsilon_{t+h}$$

- (y_1, y_2, \dots, y_t) : Sequência histórica de observações da série temporal até o tempo t .
- \hat{y}_{t+h} : Previsão do valor da série temporal no horizonte h períodos à frente, ou seja no tempo $t + h$.
- $f(y_t, y_{t-1}, \dots, y_{t-p+1})$: Função de previsão que modela a relação entre os valores atuais e passados da série. Esta função pode ser baseada em diversos modelos, como AR(p), ARIMA e até nas redes neurais.
- ϵ_{t+h} : Termo de erro ou ruído na previsão, geralmente assumido com média zero e independente das observações passadas.

3.1. Modelo SARIMAX

O SARIMAX, é um modelo derivado do original Autoregressive Integrated Moving Average (ARIMA) que é um eficiente modelo estatístico utilizado para previsão de séries temporais. Este modelo é derivado de um princípio fundamental de que os valores futuros podem ser previstos utilizando-se de características do ruído branco ϵ_t e valores do passado da série temporal [Madan and Mangipudi, 2018].

$$\Phi(L)(1-L)dy_t = \theta(L)\epsilon_t$$

3.2. Modelo Prophet

Já o Prophet, é um modelo desenvolvido por engenheiros do Facebook, ele foi projetado para lidar com dados que exibem padrões sazonais fortes, além de ser robusto a valores ausentes e mudanças de tendências. A equação geral do modelo leva em consideração componentes de tendência $g(t)$, sazonalidade $s(t)$, eventos $h(t)$ e a variação não explicada pelos componentes anteriores ou ruído branco ϵ_t [Taylor and Letham, 2017].

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

3.3. Modelo LSTM

O Long Short-Term Memory (LSTM) é um modelo avançado de rede neural recorrente (RNN), ele foi criado para resolver o problema do desaparecimento do gradiente, comum nas RNNs tradicionais, permitindo capturar informações por períodos mais longos. Ele pode ser representado pela série de equações [Zamanzadeh Darban et al., 2025].

$$\begin{aligned} f_t &= \sigma \left(W_f \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_f \right) \\ i_t &= \sigma \left(W_i \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_i \right) \\ \tilde{C}_t &= \tanh \left(W_C \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_C \right) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t &= \sigma \left(W_o \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_o \right) \\ h_t &= o_t * \tanh(C_t) \end{aligned}$$

4. Avaliação

Para avaliar, foi proposto diferentes datasets extraídos de grandes redes de computadores como ISPs e IXes. O primeiro dataset que avaliamos foi extraído de um link de trânsito com capacidade de 1 Gbps, a seguir, podemos ver os resultados de previsão para cada modelo.

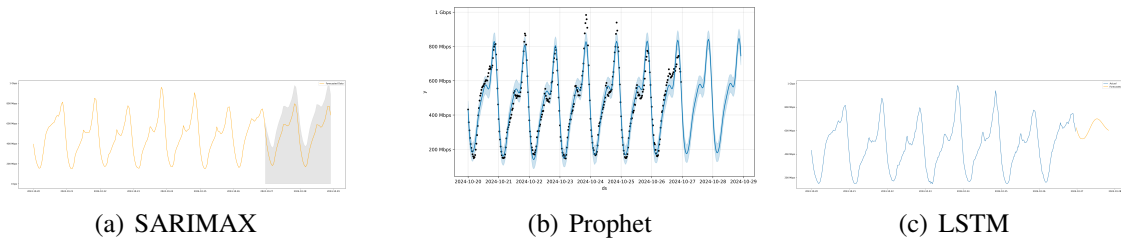


Figure 1. Comparação das previsões utilizando os três modelos

5. Resultados obtidos

Os melhores resultados de previsão obtidos foram dos modelos estatísticos, tendo em vista que o LSTM não conseguiu prever muito bem o padrão da série temporal, isso em todos os datasets que testamos. Já sob os recursos computacionais, observamos que o Prophet foi o mais "leve" em relação à tempo de treinamento e peso do modelo salvo em armazenamento secundário.

Table 1. Comparação do Modelos

Modelo	Tempo de Treinamento	Peso em Disco
LSTM	~17 segundos	11 MB
SARIMAX	~4 segundos	137 MB
Prophet	~1 segundo	140 KB

Respondendo às perguntas da seção de Objetivos da Pesquisa, temos os seguintes resultados.

- Qual algoritmo estatístico ou de inteligência artificial melhor se adequa ao cenário do tráfego da minha rede e de meus clientes?
 - S ARIMAX ou Prophet dependendo do objetivo, já que o SARIMAX foi melhor porém mais lento.
- Quais são as melhores combinações de hiper-parâmetros para meu tipo de tráfego?
 - S ARIMAX: p3, d0, q1; Prophet: padrão; LSTM: 10 épocas, 96 neuronios, 4 layers.
- O tempo para gerar o modelo versus sua eficiência de predição são adequados ao meu caso de uso levando em consideração diferentes vetores da rede?
 - O Prophet não conseguiu prever os maiores momentos de tráfego, já o SARI-MAX sim, então dependendo do contexto podemos empregar um ou o outro.

6. Possíveis melhorias e Trabalhos futuros

Como podemos observar na seção de avaliação, o modelo LSTM fez uma péssima previsão, porém imaginamos que isso seja por conta da modelagem da rede neural. Em trabalhos futuros, o autor irá desenvolver melhores redes neurais para tentar aumentar os resultados obtidos com o modelo LSTM.

Para trabalhos futuros, o autor estuda formas de correlacionar a previsão do tráfego de rede com novas features de séries temporais também extraídas dos fluxos IP de grandes redes de computadores como ISPs e IXes, com o objetivo de tentar identificar desvios de padrões, tráfego anormal e congestionamento derivado de possíveis ataques com origens globais.

7. Contexto

O seguinte trabalho e seus resultados são frutos de um projeto apresentado no final da disciplina da aula de reconhecimento de padrões do curso de Mestrado em Ciência da Computação da Universidade Estadual de Londrina (UEL) e não possui o objetivo de ser publicado como um artigo científico.

References

- Adhikari, Ratnadip and R. Agrawal (Jan. 2013). *An Introductory Study on Time series Modeling and Forecasting*. ISBN: 978-3-659-33508-2. DOI: 10.13140/2.1.2771.8084.
- Andreoletti, Davide et al. (Apr. 2019). “Network Traffic Prediction based on Diffusion Convolutional Recurrent Neural Networks”. In: *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 246–251. DOI: 10.1109/INFOCOMW.2019.8845132. URL: <https://ieeexplore.ieee.org/document/8845132>.
- Ding, Damu, Marco Savi, and Domenico Siracusa (Nov. 2022). “Tracking Normalized Network Traffic Entropy to Detect DDoS Attacks in P4”. In: *IEEE Transactions on Dependable and Secure Computing* 19.6. Conference Name: IEEE Transactions on Dependable and Secure Computing, pp. 4019–4031. ISSN: 1941-0018. DOI: 10.1109/TDSC.2021.3116345. URL: <https://ieeexplore.ieee.org/document/9552474> (visited on 11/09/2024).
- Fleischer, Antony (2017). “Using Deep Learning Techniques to Predict Network Traffic on the South African Research and Education Network”. en. In.
- Halladay, James et al. (2022). “Detection and Characterization of DDoS Attacks Using Time-Based Features”. en. In: *IEEE Access* 10, pp. 49794–49807. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3173319. URL: <https://ieeexplore.ieee.org/document/9770771/>.
- Krishnaswamy, Nandini et al. (June 2020). “Data-driven Learning to Predict WAN Network Traffic”. In: *Proceedings of the 3rd International Workshop on Systems and Network Telemetry and Analytics*. SNTA '20. New York, NY, USA: Association for Computing Machinery, pp. 11–18. ISBN: 978-1-4503-7980-9. DOI: 10.1145/3391812.3396268. URL: <https://dl.acm.org/doi/10.1145/3391812.3396268> (visited on 11/09/2024).
- Madan, Rishabh and Partha Sarathi Mangipudi (Aug. 2018). “Predicting Computer Network Traffic: A Time Series Forecasting Approach Using DWT, ARIMA and RNN”. In: *2018 Eleventh International Conference on Contemporary Computing (IC3)*. ISSN: 2572-6129, pp. 1–5. DOI: 10.1109/IC3.2018.8530608. URL: <https://ieeexplore.ieee.org/document/8530608> (visited on 11/09/2024).
- Taylor, Sean J and Benjamin Letham (Sept. 2017). *Forecasting at scale*. en. DOI: 10.7287/peerj.preprints.3190v2. URL: <https://peerj.com/preprints/3190v2> (visited on 11/09/2024).
- Troia, Sebastian et al. (Mar. 2017). “Identification of tidal-traffic patterns in metro-area mobile networks via Matrix Factorization based model”. en. In: *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. Kona, HI: IEEE, pp. 297–301. ISBN: 978-1-5090-4338-5. DOI: 10.1109/PERCOMW.2017.7917576. URL: <https://ieeexplore.ieee.org/document/7917576/>.
- Zamanzadeh Darban, Zahra et al. (Jan. 2025). “Deep Learning for Time Series Anomaly Detection: A Survey”. en. In: *ACM Computing Surveys* 57.1, pp. 1–42. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3691338. URL: <https://dl.acm.org/doi/10.1145/3691338>.