Comparando Médias Móveis com Integral de Choquet para Detectar Anomalias no Tráfego de Redes

Denner Ayres[1], Abreu Quevedo[1], Giancarlo Lucca[2], Graçaliz Dimuro[1], Bruno L. Dalmazo[1]

[1] Centro de Ciências Computacionais
Universidade Federal do Rio Grande (FURG)

[2]Universidade Católica de Pelotas (UCPel)

{dennerayres, abreu_rg, graçaliz, dalmazo}@furg.br, giancarlo.lucca@ucpel.edu.br

Abstract. *The computer network infrastructure is essential for fast and reliable access to digital resources, making it indispensable for business and daily activities. With the evident increase in continuous data flow, networks are frequently targeted by attacks. This work compares moving average models for network traffic predictions and uses the model with the lowest error to detect anomalies, comparing its performance with a data aggregation function based on the Choquet integral. The results show that the Moving Average based on the Poisson distribution outperforms the aggregation function based on the Choquet Integral with Algebraic Product.*

Resumo. *A infraestrutura de redes de computadores é essencial para o acesso rápido e confiável aos recursos digitais, sendo indispensável para negócios e atividades diárias. Com o aumento evidente do fluxo contínuo de dados, as redes são frequentemente alvo de ataques. Este trabalho compara modelos de médias móveis para previsões de tráfego de rede e utiliza o modelo com menor erro para detectar anomalias, comparando seu desempenho com uma função de agregação de dados baseada na integral de Choquet. Os resultados mostram que a Média Móvel baseada na distribuição de Poisson supera a função de agregação baseada na Integral de Choquet com Produto Algébrico.*

1. Introdução

Na sociedade moderna, a infraestrutura de redes de computadores atua como um pilar de sustentação para o acesso rápido e confiável aos recursos digitais, tornando-se indispensável tanto para os negócios quanto para as atividades diárias. O crescimento constante do número de dispositivos interconectados e sensores resulta em um fluxo contínuo de dados gerados a partir de diversas fontes e contextos. Garantir o monitoramento e a gestão eficiente dos dados nas redes é vital. Muitos esforços têm sido dedicados a garantir e melhorar a eficiência e a segurança dos dados que trafegam nas redes, sendo este um tema de extensos estudos [Dalmazo et al. 2017].

Dada a sua importância, a rede de computadores é frequentemente alvo de ataques. Alguns desses ataques deixam rastros na rede que os tornam passíveis de detecção. Nesse sentido, diversas técnicas têm sido desenvolvidas para minimizar os efeitos negativos de ações maliciosas, como, por exemplo, as abordagens de detecção de anomalias [Zeufack et al. 2021, Dalmazo et al. 2018]. Este trabalho compara modelos de médias móveis com fraca dependência histórica de dados, comumente utilizados para fazer previsões de tráfego de rede sob esse tipo de restrição. Após avaliar as médias móveis

e encontrar o modelo com menor erro na predição, utilizaremos esse modelo para detectar anomalias no tráfego de rede e comparar seu desempenho com uma função de agregação de dados baseada na integral de choquet.

O restante desse trabalho foi dividido conforme descrito. A Seção 2 apresenta os trabalhos relacionados. A arquitetura proposta é apresentada na Seção 3. A implementação e cenário de testes e resultados, respectivamente, nas Seções 4 e 5. Finalizando com as considerações finais na Seção 6.

2. Trabalhos relacionados

Nessa seção é apresentada alguns dos mais relevantes trabalhos no contexto de detecção de anomalias em redes de computadores, destacando os principais benefícios e desvantagens de cada um. Por fim, é realizada uma discussão sobre o estado da arte e os problemas em aberto.

[Yuan et al. 2023] *et al.* propõem um método fuzzy baseado em densidade é um dos comuns para detecção de anomalias. No geral, a maioria dos métodos baseados em densidade se concentra em lidar com dados certos e não considera o problema da incerteza e da imprecisão dos dados. Este artigo propõe uma detecção de anomalias baseada na densidade fuzzy-rough. Nele, a densidade fuzzy-rough é definida para descrever o grau de agregação e depois uma pontuação de anomalia é construída para caracterizar o grau de anomalia das amostras. Os resultados experimentais mostram que o método proposto bons resultados em um conjunto de dados específico.

[Schmidl et al. 2022] *et al.* implementaram mais de 70 algoritmos de detecção de anomalias de diferentes domínios e avaliaram mais de 900 conjuntos de dados de séries temporais. Os algoritmos foram selecionados de diferentes famílias de algoritmos e abordagens de detecção para representar todo o espectro das técnicas de detecção de anomalias. Esse trabalho, forneceu uma visão geral das técnicas e suas similaridades, nas quais as médias móveis apresentaram um bom desempenho com baixa dependência histórica em comparação com outras abordagens.

Entre os muitos trabalhos na literatura (os quais não podem ser listados em sua totalidade por restrições de espaço), destaca-se que os métodos fuzzy são especialmente eficazes em lidar com dados incertos e imprecisos, característica comum no tráfego de rede. Além disso, médias móveis apresentam bom desempenho com baixa complexidade devido à fraca dependência histórica. Dado esse contexto, este trabalho tem por objetivo esclarecer alguns pontos ao comparar essas duas abordagens frente à detecção de anomalias no tráfego de rede.

3. Proposta

Após um estudo prévio do melhor comportamento de janelas deslizantes, a proposta desse trabalho inicia com uma análise de erros na predição do tráfego de rede para diferentes tamanhos de janelas deslizantes, que consiste na análise de diferentes tamanhos de entradas. Após a etapa de calibração do tamanho da janela deslizante, o modelo levou em consideração os cálculos de peso entre os métodos matemáticos *Simple Moving Average* (SMA), *Weighted Moving Average* (WMA), *Exponential Moving Average* (EMA) e *Poisson Moving Average* (PMA).Conforme Figura 1, Podemos chegar às seguintes conclusões a partir desses cálculos. Todos os dados usados nesse trabalho foram coletados no dataset disponível em [Sharafaldin et al. 2018].

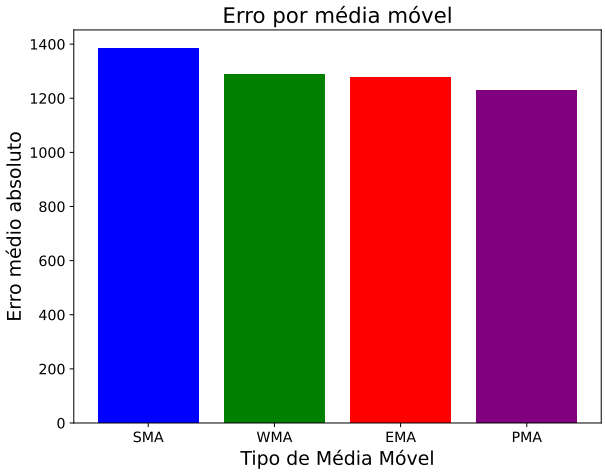

Figura 1. Comparação de erro de predição entre modelos de média móvel

Como mostra o gráfico, o PMA oferece o menor erro na predição do tráfego de rede. Complementarmente, foi avaliada a função de agregação Fuzzy baseada na Integral de Choquet, que trata os dados como dados temporais, atribuindo pesos com base em sua ordenação e tamanho. Assim, foi possível a comparação entre essas duas abordagens frente à detecção de anomalias no tráfego de rede.

No processo convencional de Choquet, a informação é agregada de acordo com a medida de importância de cada dado. Nesse trabalho usamos a Integral de Choquet com Produto Algébrico [Grabisch and Labreuche 2010], conforme a Tabela1, que leva em consideração como várias variáveis temporais interagem umas com as outras. Este método permite que os pesos sejam ajustados de forma mais precisa considerando a correlação dos dados.

Tabela 1. Tabela de exemplos de cópulas

Cópula ID	Funções
(a)	$T_M(x, y) = xy$

4. Implementação

A implementação foi projetada na linguagem Python e para lidar com a análise de séries temporais, onde a biblioteca Pandas é utilizada para processar grandes conjuntos de dados de forma eficiente. A utilização do Matplotlib facilitou a interpretação dos padrões encontrados pelos modelos preditivos, apresentando os resultados em gráficos e por fim a biblioteca Numpy para trabalhar com funções matemáticas. No modelo baseado na função de agregação de Choquet, a função de declaração de pesos é fundamental para alterar a importância de cada entrada. Com base nos pesos, a decisão é ordenada conforme o seu tamanho e em seguida são aplicadas as técnicas descritas em [Grabisch and Labreuche 2010].

O sistema de alertas implementado é essencial para a detecção de anomalias, pois identifica situações em que o valor real é significativamente diferente do valor previsto. Ao considerar as sequências de alarmes como ataques potenciais, aqui trabalhado com

uma diferença de 65 % entre o valor real e o valor predito, esta estratégia proativa fortalece o sistema para lidar com falhas e eventos imprevistos. O cálculo de erros permite quantificar e comparar a precisão dos vários modelos usados, ajudando na escolha do melhor modelo para cada situação particular de aplicação. Essa análise fornece informações úteis para o desenvolvimento contínuo de melhorias na performance e na confiabilidade dos sistemas. Todos os artefatos de software produzidos nesse trabalho podem ser acessados em: https://github.com/dennerguti/MediasMoveisChoquet.

5. Avaliação

Um conjunto organizado de procedimentos foi empregado para realizar a avaliação do sistema. Para começar os testes, decidimos usar o *Poisson Moving Average (PMA)* porque esse foi o modelo de média móvel que apresentou o melhor desempenho em nossas condições de teste. Os mesmos testes foram realizados para a função de agregação com Integral de Choquet, conforme tabelas 2 e 3, que permitiu uma avaliação completa e abrangente das soluções possíveis desse método.

A avaliação geral foi baseada no valor real que o sistema reconheceu. Esse valor demonstra a capacidade do sistema de identificar e processar dados com precisão em diferentes situações. Este critério consolidou os resultados dos testes em uma avaliação significativa e forneceu uma medida objetiva do desempenho geral do sistema. Para avaliar a eficiência desses novos modelos, a Matriz de Confusão será usada. Isso permitirá uma análise minuciosa das taxas de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos, o que contribuirá para uma comparação de desempenho dos modelos avaliados.

Tabela 2. Matriz de confusão PMA.

		Classe prevista	
		0	1
Classe esperada	0	VN: 7183	FP: 9535
	1	FN: 11816	VP: 8378

Tabela 3. Matriz de confusão Choquet.

		Classe prevista	
		0	1
Classe esperada	0	VN: 7183	FP: 9535
	1	FN: 12826	VP: 8001

Aplicando então as seguintes métricas $TVP = \frac{VP}{VP+FN}$ e $TVN = \frac{VN}{VN+FP}$, podemos observar a Tabela2 que apresenta a matriz confusão da metodologia PMA. A partir dela obtém-se os valores de 0,43 e 0,52 para TVP e TVN respectivamente. O que indica 43% de acerto na detecção da classe 0 e 52% de acerto na detecção da classe 1. Na Tabela 3, observamos que a metodologia de Choquet teve os valores de 0,43 e 0,62 para TVP e TVN respectivamente. Indicando 43% de detecção para a classe 0 e 62% para a classe 1.

6. Conclusão

Os dados analisados nas tabelas 4 e 5 mostram, que o modelo já proposto, aqui chamado de modelo PMA, continua a fornecer os melhores resultados de predição. Em

Tabela 4. Tabela de Alarmes

Tipo de alarme	Quantidade
Alarme Geral	16718
Alarme PMA	20194
Alarme Choquet	20827

Tabela 5. Tabela de Ataques

Tipo de ataque	Quantidade
Ataque Geral	7183
Ataque PMA	11816
Ataque Choquet	12826

comparação com outros métodos, este modelo demonstrou ser mais preciso e sensível a mudanças. Mas a metodologia com Integral de Choquet demonstrou um potencial significativo em suas previsões, com valores que se aproximam muito do cenário ideal.

O trabalho ainda está em andamento, como trabalhos futuros, vamos usar uma metodologia de treshold dinâmico para definir anomalias, filtrar os ruídos, usar outros modelos especializados da Integral de Sugeno e Choquet, assim como verificar a generalização da solução proposta através de testes com outros datasets.

Acknowledgment

The authors would like to thank FAPERGS/CNPq (23/2551-0000773-8); UE iTec/FURG (iTec-80) and Cnpq (407206/2023-0, 304118/2023-0) for partially supporting this work.

Referências

Dalmazo, B. L., Vilela, J. P., and Curado, M. (2017). Performance analysis of network traffic predictors in the cloud. *Journal of Network and Systems Management*, 25:290–320.

Dalmazo, B. L., Vilela, J. P., and Curado, M. (2018). Triple-similarity mechanism for alarm management in the cloud. *Computers & Security*, 78:33–42.

Grabisch, M. and Labreuche, C. (2010). A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. *Annals of Operations Research*, 175(1):247–286.

Schmidl, S., Wenig, P., and Papenbrock, T. (2022). Anomaly detection in time series: a comprehensive evaluation. *Proc. VLDB Endow.*, 15(9):1779–1797.

Sharafaldin, I., Lashkari, A. H., and Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *4th International Conference on Information Systems Security and Privacy (ICISSP)*, Portugal.

Yuan, Z., Chen, B., Liu, J., Chen, H., Peng, D., and Li, P. (2023). Anomaly detection based on weighted fuzzy-rough density. *Applied Soft Computing*, 134:109995.

Zeufack, V., Kim, D., Seo, D., and Lee, A. (2021). An unsupervised anomaly detection framework for detecting anomalies in real time through network system's log files analysis. *High-Confidence Computing*, 1(2):100030.