

ARTIGO COMPLETO/FULL PAPER

O Parâmetro α na generalização da Integral de Choquet para Previsão de Tráfego de Rede

The α Parameter in the Generalized Choquet Integral for Network Traffic Prediction.

Abreu Quevedo • ✉ abreurg@furg.br

Universidade Federal do Rio Grande (FURG)

Denner Ayres • ✉ dennerayres@furg.br

Universidade Federal do Rio Grande (FURG)

Graçaliz Dimuro • ✉ gracaliz@furg.br

Universidade Federal do Rio Grande (FURG)

Giancarlo Lucca • ✉ giancarlo.lucca@ucpel.edu.br

Universidade Católica de Pelotas (UCPel)

André Riker • ✉ ariker@ufpa.br

Universidade Federal do Pará (UFPA)

Bruno L. Dalmazo • ✉ dalmazo@furg.br

Universidade Federal do Rio Grande (FURG)

RESUMO. O tráfego de rede é essencial para o funcionamento do cotidiano moderno, e prever e mensurar esse tráfego é um recurso estratégico valioso, exigindo conhecimento profundo de algoritmos e tecnologia. Diversos modelos, como fuzzy c-means e regressão, tentam resolver esse problema, mas ainda não atingem o desempenho ideal. Este trabalho propõe um modelo de agregação baseado na variação do parâmetro α na integral de Choquet, buscando mensurar os erros associados a cada valor de α para identificar o melhor método e parâmetro para otimização das equações.

ABSTRACT. Network traffic is crucial to the functioning of modern life, and predicting and measuring this traffic is a valuable strategic asset that requires deep knowledge of algorithms and technology. Several models, such as fuzzy c-means and regression, attempt to solve this issue but still do not achieve optimal performance. This paper proposes an aggregation model based on the variation of the α parameter in the Choquet integral, aiming to measure the errors associated with each α value to identify the best method and parameter for optimizing the equations.

PALAVRAS-CHAVE: Modelos de busca • Integral de Choquet • Predição de tráfego de redes

KEYWORDS: Search models • Choquet integral • Network traffic prediction

1 Introdução

Na sociedade atual em que estamos inseridos, é evidente que estamos altamente dependentes da infraestrutura de rede de computadores, visto que é um modelo que possibilita o acesso rápido e confiável a recursos digitais, tornando-se essencial para as atividades cotidianas e empresariais [1].

O tráfego de informações na rede precisa ser assegurado, monitorado e bem administrado para garantir tanto o fluxo quanto a segurança dos dados. Com o aumento do volume e a constante variação dessas informações, métodos que aprimorem a eficiência e a segurança [2, 3] se tornam cada vez mais essenciais. A predição de tráfego de rede é um desses métodos, que pode melhorar a detecção de anomalias, garantir uma distribuição eficiente do tráfego e reduzir falhas. Uma abordagem promissora para esse fim é a integral de Choquet difusa, que proporciona flexibilidade na mo-

delagem de preferências e na tomada de decisões em ambientes incertos, sendo uma via de discussão para soluções para a predição de tráfego de rede.

2 Trabalhos Relacionados

Em [4] é proposto a utilização de um modelo híbrido para prever o tráfego de rede e melhorar a qualidade do serviço “QoS”, para isso, são utilizados modelos avançados de previsão de séries temporais, com o fuzzy c-means sendo empregado os dados da rede, e assim, melhorando a série temporal existente, sendo possível obter o “Clustering granule”, a metodologia utilizada pelo artigo, é do estudo utilizando-se de dados diferentes, e também, comparando o modelo proposto com modelos alternativos apresentados durante a apresentação do mesmo. A conclusão deste artigo foi de que os resultados de predição do modelo proposto são mais satisfatórios, sendo fornecido melhores resultados com

a I.A e aumentando de forma efetiva a precisão da previsão do tráfego de rede.

Em [5] temos o objetivo de obter um método de amostragem adaptativa baseado em Regression Model e lógica Fuzzy, sendo esta desenvolvida para redes sem fio, com a intenção de reduzir a quantidade de dados transmitidos por pacotes. O método utilizado, foi feito através da criação de um projeto de uma scalable network, utilizando alguns pacotes de simulação de rede, sendo o projeto criado, seguindo uma rede hierárquica recomendada, sendo assim utilizado essa rede para serem feitos testes. Os tipos de tráfego utilizados para os testes foram os de streaming de vídeo, VoIP, HTTP e FTP. A conclusão dos estudos foi de que comparando o modelo adaptativo apresentado no artigo, com os modelos não adaptativos, o que trouxe um melhor resultado, e uma maior precisão foi o modelo adaptativo.

Em [6] se tem como objetivo solucionar o problema da baixa precisão de detecção de ataques, através da distribuição desigual do tráfego de rede. Para resolver este problema é proposto, uma nova abordagem chamada “the fuzzy entropy weighted natural nearest neighbor (FEW-NNN) method” para aumentar a precisão e a eficiência da detecção de ataques de tráfego de rede baseada em fluxo. Para a realização destes testes, foram utilizados alguns datasets como amostras para o estudo, sendo estes o KDD99 e a CIC-IDS-2017, diante disto, foi observado que o método “FEW-NNN” melhora de forma agradável a precisão e a eficiência da detecção de ataques de tráfego de rede baseada em fluxo, tendo uma perspectiva boa no campo detecção de intrusão de rede.

Em [7] é apresentado uma análise de desempenho de preditores de tráfego de rede na nuvem, com foco em selecionar modelos de previsão adequados para ambientes em nuvem. O trabalho propõe um mecanismo de análise padronizado para avaliar os melhores modelos de previsão candidatos em termos de precisão, dependência histórica, tempo de execução e eficiência computacional. A conclusão final deste trabalho é de que A partir da observação dos resultados do estudo de caso do Dropbox, pode-se perceber que todas as previsões baseadas na análise local apresentam uma melhora considerável usando o algoritmo de tamanho de Janela dinâmica (DyWiSA), facilitando também a previsão de tráfego online devido à sua curta dependência de dados históricos. O Poisson Moving Average, apesar de apresentar bons resultados, manteve a mesma complexidade computacional, que a dos seus concorrentes baseados em análise local. Seguindo com o ARIMA que

foi utilizado no dataset "Data Centre", foi obtido uma vantagem significativa em relação aos outros preditores, todavia, essa vantagem foi obtida através de uma grande complexidade computacional e tempo gasto. O Poisson Moving Average, o qual é mais atrativo por seu baixo custo de complexidade computacional, se mostrou mais adequado para ambientes dinâmicos na nuvem, do que os outros modelos.

3 Proposta

A proposta deste trabalho é realizar uma comparação entre os diferentes resultados obtidos com a variação do parâmetro α , utilizado na generalização da integral de Choquet, com o objetivo de prever as tendências do tráfego de rede de forma mais assertiva.

3.1 Modelo conceitual

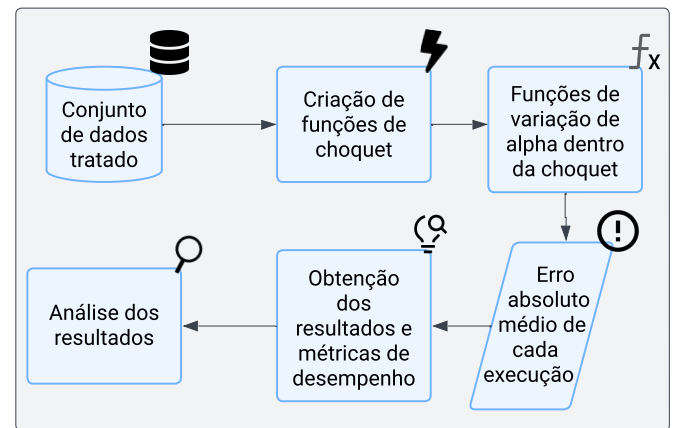


Figura 1. Diagrama do modelo conceitual.

O modelo conceitual apresentado na Figura 4 ilustra o fluxo de trabalho para análise do α e previsão de dados utilizando funções de Choquet.

1. **Dados e funções:** Preparação e limpeza dos dados para análise. Inclui assim, a remoção de valores atípicos, preenchimento de valores ausentes e normalização dos dados dentre outras maneiras de visar a adequação da base de dados para as próximas etapas.
2. **Criação de funções de Choquet:** Implementação de funções baseadas na integral de Choquet para agregação e previsão dos dados de rede.
3. **Funções de variação de α dentro da Choquet:** Criação dos métodos de busca para variação do parâmetro α no contexto da integral de Choquet.
4. **Execução das funções nos dados:** Aplicação das funções de Choquet e da busca do α na lista de tráfego de rede.

5. Obtenção e análise dos Resultados e Métricas de Desempenho: Realiza-se a compilação dos resultados, junto de tabelas e gráficos, para pôr fim realizar a análise e interpretação de desempenho de cada α , bem como a eficiência da metodologia de busca.

3.2 Dataset utilizado

Neste estudo, foi utilizada uma versão tratada do dataset CIC-DDoS2019, criado pelo Canadian Institute for Cybersecurity. O dataset oferece um volume significativo de dados rotulados, essenciais para o treinamento de algoritmos de aprendizado de máquina, incluindo uma variedade de ataques DDoS. A principal característica deste trabalho é o uso de 29.404 valores de tráfego de rede coletados em um período de 24 horas.

3.3 Algoritmos para variação do α

Para a implementação da realização das comparações e testes em cima dos algoritmos, foram executadas duas soluções, a primeira foi a busca por força bruta e como segunda metodologia foi utilizada a busca binária.

O ambiente utilizado para execução dos algoritmos foi uma máquina virtual do Google Colab. A máquina utilizada comporta um processador Intel(R) Xeon(R) CPU @ 2.20GHz 1 núcleo e 2 threads, com 13gb de memória ram, sendo a versão 3.10.12 11-09-2024, GCC 11.4.0. Junto ao Colab foi reproduzido o algoritmo na linguagem de programação Python 3, em sua versão 3.10.12.

3.3.1 Método de força bruta

O algoritmo de busca por força bruta, tem como objetivo otimizar um parâmetro denominado α , através da minimização de um erro associado ao valor estimado e ao verdadeiro valor. Para isso, ele percorre um vetor de α testando todas as opções e testando seus erros, guardando sempre o com menor erro associado.

3.3.2 Método de busca binária

O algoritmo de busca binária tem como objetivo otimizar um parâmetro denominado α , através da minimização de um erro associado ao valor estimado e ao verdadeiro valor. Diferentemente da busca por força bruta, que testa todas as opções, a busca binária opera em um vetor ordenado, dividindo repetidamente o intervalo de busca pela metade. O algoritmo começa avaliando o ponto médio do intervalo, comparando o erro associado a esse valor com o erro de outros valores em potencial, dependendo do resultado dessa comparação, o algoritmo ajusta o intervalo de busca para a metade superior ou inferior, continuando esse processo até encontrar o valor de α que satisfaça um erro baixo.

3.4 Funções de agregação e integral de Choquet

A integral de Choquet [8] é usada para agregação flexível de dados com base na importância de cada critério, definida por uma medida difusa $m : 2^N \rightarrow [0, 1]$ [9].

$$C_m(x) = \sum_{i=1}^n (x_{\sigma(i)} - x_{\sigma(i-1)}) m(A_{(i)}),$$

onde $(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ é uma permutação dos dados de entrada em ordem crescente e $A_{(i)} = \{(i), \dots, (n)\}$ é o subconjunto dos índices das maiores componentes.

Copulas [10] são funções matemáticas, $C : [0, 1]^2 \rightarrow [0, 1]$, usadas para conectar distribuições bidimensionais às suas marginais unidimensionais. Elas são importantes para espaços métricos probabilísticos e têm aplicações em estatística.

A integral de Choquet pode ser expandida, substituindo os operadores de produto por métodos de cópula [11]

$$C_m(x) = \sum_{i=1}^n (x_{(i)} \cdot m(A_{(i)}) - x_{(i-1)} \cdot m(A_{(i)})),$$

As funções de cópula consideradas estão na Tabela 1, e a notação da coluna 1 será usada ao longo do texto [12].

4 Avaliação

Para ser mostrado a viabilidade desta técnica proposta por este trabalho, esta seção analisa a influência da variação do parâmetro α na integral de Choquet e o impacto do mesmo, condizendo com o erro de previsão resultado pelo mesmo.

4.1 Caso de estudo 1 - Busca por força bruta

Na abordagem inicial desta investigação, foi utilizada uma técnica conhecida como "busca por força bruta". Essa abordagem é caracterizada pela sua simplicidade e rigidez na exploração de todas as possíveis soluções de um problema. No contexto do nosso estudo, isso implica em percorrer de maneira exaustiva uma lista composta por uma série de valores α , sendo que cada uma dessas listas, correspondem a uma faixa dos parâmetros suportados por um conjunto de equações específicas da integral de Choquet.

A Tabela 2 apresenta os resultados obtidos pelo algoritmo de força bruta. É possível visualizar os melhores resultados encontrados para cada uma das equações de Choquet, incluindo o valor do parâmetro α que resultou no menor erro médio absoluto.

Tabela 1. Tabela da generalização das Choquet usadas neste trabalho

Copula ID	Funções	Propriedades
(a)	$C_\alpha(x, y) = xy[1 + \alpha(1 - x)(1 - y)]$	$-1 \leq \alpha \leq 1 (\alpha \neq 0)$
(b)	$C_\alpha(x, y) = \frac{1}{1+\alpha} \text{Max}[x + y - 1 + \alpha - \alpha x - y , 0]$	$0 < \alpha < 1$
(c)	$C_\alpha = (1 - \alpha)W + \alpha \text{Min}$	$0 < \alpha < 1$

Equação	Melhor α	Média dos erros	Tempo
Choquet a	-1.0	1068.49	49 segundos
Choquet b	0.6	1120.41	49 segundos
Choquet c	0.23	1332.44	132 segundos
Total			3 min 50 segundos

Tabela 2. Tabela de resultados de α para busca de força bruta

4.2 Caso de estudo 2 - Busca binária

O segundo caso de estudo (apresentado na Tabela 3), temos a introdução da "busca binária". Com este algoritmo, foi realizado em Python, experimentos visando também encontrar os valores ótimos de α para realizar previsões junto as equações de Choquet.

Com a utilização da busca binária, foram obtidos resultados satisfatórios, sendo possível visualizar na tabela 3 estes resultados. Levando em consideração que a busca binária é conhecida por sua eficácia na otimização de parâmetros, onde se procura refinamento em um intervalo limitado, os resultados obtidos foram fundamentais para entender como se comporta o α contexto do nosso estudo.

4.3 Discussão dos Resultados

Diante da análise do resultado das tabelas e dos casos de estudos, observa-se que foi realizada a avaliação de dois métodos distintos de busca, sendo elas por força bruta e busca binária para determinar o impacto do parâmetro α na integral de Choquet.

4.3.1 Métodos

Força bruta foi o método de percorrer o alpha ideal mais lento, levando 3 minutos e 50 segundos, enquanto a busca binária reduziu esse tempo para aproximadamente 1 minuto e 45 segundos. A busca binária mostrou-se a mais eficiente, oferecendo alta precisão e tempo de execução otimizado perante suas execuções.

4.3.2 α e sua variação

Os valores de α obtidos variaram entre os métodos, indicando a sensibilidade do parâmetro nas diferentes abordagens. Relacionando o parâmetro com a melhora da média dos erros médios absolutos, foi possível observar que ele varia em cada equação da tabela 1. É

possível ver com mais detalhes, observando o gráfico da Figura 3.

4.3.3 Resumo das discussões

Este estudo mostrou que a escolha do método de busca do parâmetro α é fundamental para a precisão e eficiência na aplicação da integral de Choquet para previsão de tráfego de redes. Com isso métodos mais rápidos como o da busca binária podem ser preferíveis em cenários com um tamanho muito grande de dados, e um curto período de tempo, como o realizado por este trabalho. Além disso foi afirmada a importância de escolher um α apropriado para a equação utilizada, a fim de se obter resultados melhores e mais assertivos, observando o impacto significativo do α diante a utilização da Choquet.

Equação	Melhor α	Média dos erros	Tempo
Choquet a	-0.98	1069.14	19 segundos
Choquet b	0.49	1121.34	23 segundos
Choquet c	0.25	1332.66	63 segundos
Total			1 min 45 segundos

Tabela 3. Tabela de resultados de α para busca binária

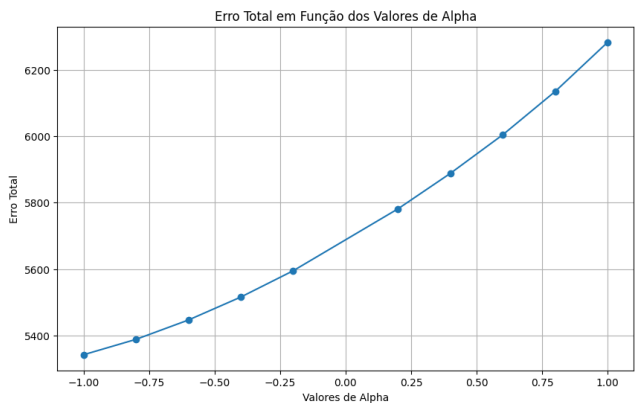


Figura 2. Gráfico da integral de Choquet a.

5 Conclusão

Este estudo apresentou a relevância do parâmetro α nas equações da integral de Choquet no contexto de previsão de tráfego de redes, assim como métodos eficientes para encontrar os parâmetros ideais de cada equação.

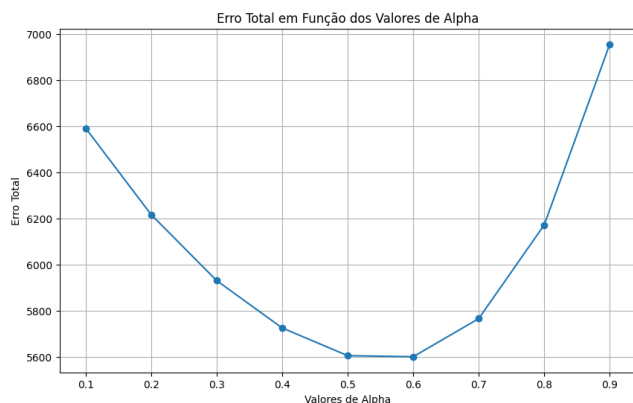


Figura 3. Gráfico da integral de Choquet b.

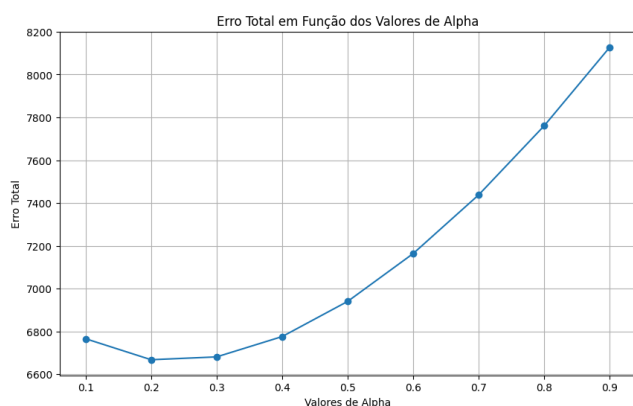


Figura 4. Gráfico da integral de Choquet c.

Demonstrando que a seleção dos α impacta de forma positiva na melhora do desempenho das previsões, bem como a metodologia de seleção.

Por fim, em um trabalho futuro pretende-se usar os conhecimentos desse trabalho para dar suporte a um Sistema de Detecção de intrusão real. Espera-se a partir de uma estimativa do α e uma escolha correta de atributos, seja possível atingir um *trade-off* favorável, ou seja, ter resultados eficientes, ao mesmo tempo que não use todos os recursos disponíveis, economizando memória e tempo de processamento.

Declarações complementares

Financiamento

Os autores expressam seu sincero agradecimento à FAPERGS (24/2551-0001396-2) e à FAPERGS/CNPq (23/2551-0000126-8; 23/2551-0000773-8) pelo apoio financeiro essencial para a realização deste trabalho.

Referências

- Cloud, H. The nist definition of cloud computing. *National institute of science and technology, special publication*, v. 800, n. 2011, p. 145, 2011.

- Dalmazo, B. L.; Vilela, J. P.; Curado, M. Triple-Similarity Mechanism for alarm management in the cloud. *Computers & Security*, v. 78, p. 33–42, 2018. ISSN 0167-4048. DOI: <https://doi.org/10.1016/j.cose.2018.05.016>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167404818306515>.
- Leite, L. et al. Federated Learning under Attack: Improving Gradient Inversion for Batch of Images. In: ANAIS do XXIV Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais. São José dos Campos/SP: SBC, 2024. P. 794–800. DOI: [10.5753/sbseg.2024.241680](https://sol.sbc.org.br/index.php/sbseg/article/view/30070). Disponível em: <https://sol.sbc.org.br/index.php/sbseg/article/view/30070>.
- Aldhyani, T. H. H. et al. Intelligent Hybrid Model to Enhance Time Series Models for Predicting Network Traffic. *IEEE Access*, v. 8, p. 130431–130451, 2020. DOI: [10.1109/ACCESS.2020.3009169](https://doi.org/10.1109/ACCESS.2020.3009169).
- Salama, A.; Saatchi, R.; Burke, D. Adaptive Sampling Technique for Computer Network Traffic Parameters Using a Combination of Fuzzy System and Regression Model. In: 2017 Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI). 2017. P. 206–211. DOI: [10.1109/MCSI.2017.43](https://doi.org/10.1109/MCSI.2017.43).
- Chen, L. et al. FEW-NNN: A fuzzy entropy weighted natural nearest neighbor method for flow-based network traffic attack detection. *China Communications*, v. 17, n. 5, p. 151–167, 2020. DOI: [10.23919/JCC.2020.05.013](https://doi.org/10.23919/JCC.2020.05.013).
- Dalmazo, B. L.; Vilela, J. P.; Curado, M. Performance analysis of network traffic predictors in the cloud. *Journal of Network and Systems Management*, Springer, v. 25, p. 290–320, 2017.
- Mayor, G.; Trillas, E. On the representation of some aggregation functions. In: *Fuzzy sets and Systems*, Elsevier, v. 64, n. 1, p. 73–86, 1994.
- Murofushi, T.; Sugeno, M.; Machida, M. Non-monotonic fuzzy measures and the Choquet integral. *Fuzzy sets and Systems*, Elsevier, v. 64, n. 1, p. 73–86, 1994.
- Alsina, C.; Schweizer, B.; Frank, M. J. *Associative functions: triangular norms and copulas*. World Scientific, 2006.
- Lucca, G. et al. A proposal for tuning the α parameter in $C_\alpha C$ -integrals for application in fuzzy rule-based classification systems. *Natural Computing*, v. 19, n. 3, p. 533–546, 2020. ISSN 1572-9796. DOI: [10.1007/s11047-018-9678-x](https://doi.org/10.1007/s11047-018-9678-x). Disponível em: <https://doi.org/10.1007/s11047-018-9678-x>.
- Lucca, G. et al. A Proposal for Tuning the Alpha Parameter in a Copula Function Applied in Fuzzy Rule-Based Classification Systems. In: 2016 5th Brazilian Conference on Intelligent Systems (BRACIS). 2016. P. 367–372. DOI: [10.1109/BRACIS.2016.073](https://doi.org/10.1109/BRACIS.2016.073).