

Alocação de Recursos em Redes Sem Fio de Grande Porte por meio de Meta-Heurísticas e Predição da Carga de Usuários

Lucas R. Frank¹, Lorenzo Carnevale², Antonino Galletta², Massimo Villari²,
Alex B. Vieira¹, Edelberto Franco Silva¹

¹ Departamento de Ciência da Computação,
Universidade Federal de Juiz de Fora, Juiz de Fora, Brasil

{lucasrodrigues, alex.borges, edelberto}@ice.ufjf.br

²Department of Mathematical and Computer Science, Physical Science and Earth Sciences,
University of Messina, Messina, Italy

{lcarnevale, angalletta, mvillari}@unime.it

Abstract. *With the significant increase in mobile users, the demand for resource management has become essential. This management must aim to meet the signal coverage, but mainly to maintain your Service Level Agreement, regardless of the number of connected users. Thus, we propose the use of four prediction models applied to the number of connected users in a wireless network. From these predictions, network resources can be allocated appropriately. We investigate the use of Particle Swarm Optimization and Genetic Algorithm to hyperparameterize a Multilayer Perceptron and a Decision Tree. We evaluate our proposal using real wireless network data with more than 20,000 users. As a result, we obtained an average accuracy of 94.80%, considerably improving the network resource utilization and meeting a service level agreement of 95%.*

Resumo. *Com o aumento significativo de usuários móveis, a busca pelo gerenciamento de recursos tornou-se essencial. Essa gerência deve visar atender à cobertura do sinal, mas, principalmente, manter o Acordo de Nível de Serviço desejado, independentemente do número de usuários conectados. Assim, propomos a utilização de quatro modelos de predição aplicado ao número de usuários conectados em uma rede sem fio. A partir dessas previsões, os recursos de rede podem ser alocados adequadamente. Investigamos o uso de Otimização de Enxame de Partículas e Algoritmo Genético para hiperparametrizar um Perceptron Multicamadas e uma Árvore de Decisão. Avaliamos nossa proposta utilizando dados reais de rede sem fio com mais de 20 mil usuários. Como resultado, obtivemos uma precisão média de 94,80%, melhorando consideravelmente a utilização de recursos da rede e atendendo um nível de acordo de serviço de 95%.*

1. Introdução

Grandes mudanças aconteceram nas últimas décadas em redes de computadores, desde o incremento no número de dispositivos pessoais sem fio até a velocidade e cobertura ofertadas para acesso a Internet. Destacam-se as redes locais sem fio (WLAN), presentes em praticamente todos os lugares, como aeroportos, hospitais e escolas. Uma rede sem fio

infraestruturada depende de pontos de acesso (AP), sua distribuição geográfica e também da intensidade do sinal emitido. Porém, os recursos de rede são limitados e a relação entre cobertura e economia de recursos um desafio.

Além disso, quanto mais dispositivos conectados, maior a necessidade de cobertura. Maiores também são as demandas por recursos como largura de banda disponível em uma determinada área [Prasad and Agrawal 2010] e capacidade para atendimento de usuários simultâneos [Medeiros et al. 2019]. Claramente, os recursos da rede sem fio precisam ser gerenciados da melhor maneira possível evitando a escassez e mantendo a qualidade do serviço (QoS) e a qualidade da experiência (QoE) dos usuários.

Nesse sentido, é interessante considerar as tomadas de decisões de forma inteligente, com foco nos usuários da rede móvel. Por exemplo, com o aumento do número de usuários móveis e da cobertura das WLANs, a associação de usuários a APs durante sua movimentação é frequente. Assim, é essencial entender os padrões de mobilidade e dessas associações para que soluções viáveis possam ser planejadas. Além disso, redes sem fio de larga escala geram grandes massas de dados, onde uma semana de dados pode ter mais de 500.000 usuários associados aos APs, como é o caso do ambiente avaliado em nosso estudo. Devido a mobilidade de usuários desse ambiente para com o resto da cidade, a computação urbana auxilia nesse estudo para que o serviço possa atender todos os usuários com qualidade. Assim, o número de usuários conectados influencia diretamente na necessidade de melhorar a alocação de largura de banda disponível, deste modo garantindo uma melhor QoE, e mantendo o acordo de nível de serviço (SLA) contratado [Datsika et al. 2018, Ibrahim et al. 2016].

Uma das tomadas de decisões de forma inteligente, centradas no usuário, é a predição do número de usuários conectados. Isso possibilita melhorar o gerenciamento de recursos da rede sem fio e fornecer uma melhor QoE para eles. De fato, com a predição de tráfego, é possível gerenciar recursos como: largura de banda alocada, ligar ou desligar APs de acordo com a quantidade de usuários previsto, entre outras funções. Assim, é possível gerenciar os dispositivos de rede (*e.g.*, APs) a fim de reduzir o consumo de energia, por exemplo, balanceando¹ a associação de usuários entre os APs. Em suma, pode-se evitar o uso indevido de APs quando não é necessário que todos estejam ativos, mantendo o SLA esperado pelos usuários da rede.

Neste trabalho, avaliamos 4 modelos preditivos com dados de uma rede real com 20 mil usuários. Através desses modelos foram avaliadas as demandas dos AP (números de conexões neles). Como o uso de metaheurísticas para otimização de hiperparâmetros tem se tornado comum [Bacanin et al. 2020], logo, adotamos duas metaheurísticas a fim de otimizar os hiperparâmetros dos modelos, *i.e.*, um *Particle Swarm Optimizer* (PSO) e um Algoritmo Genético (GA) que são algoritmos simples. Em seguida, treinamos um *Multilayer Perceptron* (MLP) e uma *Decision Tree* (DT) para prever o número de usuários conectados em um determinado AP. Também aplicamos os modelos para alocar apenas os APs necessários. Nossos resultados mostram que, em média, nosso melhor modelo indica um coeficiente de determinação de 0,9434 e uma acurácia de 94.80% na alocação de APs. Com esse desempenho, nosso modelo consegue alocar de forma eficiente a quantidade necessária de APs. Desta forma, auxiliando na economia de recursos, sejam estes da

¹Forçando o *handoff* - a associação do usuário em um outro AP próximo a ele com a intenção de desligar aquele AP original que apresenta baixa carga de usuários.

rede, como também o gasto para manter esses equipamentos ligados.

Nossas contribuições são: (i) a utilização de duas metaheurísticas para otimizar os hiperparâmetros dos modelos e prever o número de usuários conectados em uma rede sem fio; (ii) a avaliação da proposta em um conjunto de dados real obtido de uma rede sem fio de grande porte; e (iii) análise do impacto desta proposta no cenário de gestão da rede.

O resto do artigo está organizado da seguinte forma. A Seção 2 descreve os trabalhos relacionados. Já como os dados são pré-processados é descrito na Seção 3. A implementação da abordagem proposta está na Seção 4, enquanto os experimentos são discutidos na Seção 5. Conclusões e nossas direções futuras estão resumidas na Seção 6.

2. Trabalhos Relacionados

O aumento explosivo do volume de dados e dispositivos móveis conectados trouxe desafios significativos para o gerenciamento e otimização do tráfego de rede [Fu et al. 2018]. Uma das formas de lidar com esse desafio é encontrar uma forma autônoma para gerenciar o congestionamento da rede. Uma solução interessante é aplicar os conceitos de *Software-Defined Networking* (SDN), que isolam naturalmente os planos de controle e dados, a fim de oferecer suporte ao provisionamento rápido de serviços, e minimizar as restrições de *hardware* [Fu et al. 2018]. Por exemplo, a proposta de balanceamento de carga apresentado por [Vieira et al. 2021], determina, com base nas informações de carga da rede e topologia, quais roteadores devem ser mantidos ligados ou não, melhorando assim a gestão dos recursos.

Aplicar Inteligência Artificial (IA) é uma forma promissora para gerir a carga de redes sem fio [Chih-Lin et al. 2017]. Tanto redes neurais (NN) quanto modelos baseado em árvore de decisão fornecem melhores resultados que os modelos *Naive Bayes* (NB), *Support Vector Machines* (SVM), *Random Forest* (RF), *k Nearest Neighbors* (kNN) em termos de precisão e predição, conforme [Casas et al. 2018]. Em particular, os modelos baseados em árvore de decisão representam uma solução atraente para análise de rede sem fio, principalmente, devido ao seu baixo custo computacional. Nesta linha [Li et al. 2020b] propõem o uso de uma Rede Neural de *Backpropagation* (BPNN) após uma fase de pré-processamento otimizada aplicando *denoising wavelet*. O erro médio absoluto (MAE) alcançado foi menor do que o modelo *Autoregressive-moving-average* (ARMA). Por sua vez, o Aprendizado de Máquina (*Machine Learning* ou ML) oferece suporte mesmo quando é necessário caracterizar ou inferir os estados do protocolo da camada de aplicação por meio de fluxos diversos de tráfego de rede sem fio intercalados. Como pode ser visto em [Ford et al. 2018]. Os trabalhos citados deixam claro os benefícios da utilização de técnicas de IA como auxiliar na gestão da rede de comunicação.

Outros resultados relevantes também são encontrados na literatura ao aplicar-se *Deep Learning* na predição de tráfego de rede sem fio. Em [Li et al. 2020a] foi proposta uma rede convolucional espaço-temporal (LA-ResNet), combinando rede residual, Rede Neural Recorrente (RNN) e um mecanismo de atenção. A rede residual é usada para modelar as características espaciais dos dados de sequência. Em contraste, a RNN e um mecanismo de atenção são combinados para resolver a modelagem espaço-temporal e prever o tráfego de rede sem fio. O RNN também foi aplicado junto com a decomposição do tensor Bayesiano Gaussiano para construir um modelo de predição de

tráfego de rede sem fio. O algoritmo, chamado BGCP-RNN-ReLU, pode prever efetivamente as mudanças no tráfego de rede tanto *upstream*, quanto *downstream*, para um futuro próximo [Deng et al. 2021]. Por si só, destaca-se que uma RNN simples pode obter melhor desempenho de predição do que outras NN complexas. Mas, ainda assim, a RNN *Long-Short Term Memory* (LSTM) foi comparada com o modelo *Gated Recurrent Unit* (GRU) para prever o tráfego de rede sem fio e os dados de utilização do canal [Sone et al. 2021]. Como resultado, concluiu-se que modelos de ML são adequados para previsões de média escala. Já para prever dados de larga escala, *HoltWinters* e a Média Móvel Integrada Sazonal Autorregressiva (SARIMA) se destacam, principalmente por conta da sua insensibilidade a valores discrepantes do conjunto de dados.

Em [Saikhu et al. 2020] foi proposta uma solução para o problema de séries temporais multivariadas (MTS) também para prever o tráfego de rede sem fio. Em primeiro lugar, foi construída uma abordagem de correlação espaço-temporal por meio do método *Detrended Partial Cross Correlation Analysis* (DPCCA) para extração de características, e após isso, aplicado um modelo RNN para prever o tráfego de redes sem fio. A abordagem oferece melhor desempenho do que um modelo *Generalized Spatio-Temporal Auto-Regressive* (GSTAR), uma vez que não envolve preditores externos. Por fim, uma BPNN também foi avaliada como arquitetura básica de um modelo de predição de tráfego de rede sem fio. Na qual uma análise, aplicando o método *wavelet* de multi-resolução, compreendeu o pré-processamento de dados, e um algoritmo aprimorado *Quantum Particle Swarm Optimization* (ALQPSO) como otimização [Yu et al. 2020]. Comparado com uma Rede Neural BP combinado à otimização por Algoritmo Genético (GA-BPNN) e uma Rede Neural BP combinado à otimização por Enxame de Partículas Quânticas (QPSO-BPNN), seu valor de erro é menor e, portanto, apresenta uma predição mais precisa.

Diferentemente dos trabalhos citados, propomos uma comparação somente entre modelos leves em termos computacionais e de maior interpretação, sendo eles: uma rede neural do tipo MLP e DT. Ambos modelos foram otimizados por metaheurísticas, e, por fim, aplicados na predição do número de usuários conectados na rede sem fio. O PSO e o GA foram as metaheurísticas escolhidas, criando as combinações: PSO-MLP, GA-MLP, PSO-DT e GA-DT. Uma outra lacuna – além da proposta de modelos mais simples – percebida no estado da arte é a avaliação dos modelos de predição dentro de um cenário de dados reais para verificação da economia na alocação de largura de banda disponível e do espectro sem fio, de forma que mantenha um SLA contratado, mas reduza o número de recursos/APs necessários para isso.

3. Caracterização de dados

Inicialmente, realizamos a caracterização dos dados da rede sem fio do campus da Universidade Federal de Juiz de Fora. Os dados utilizados para análise foram coletados no período de junho a setembro de 2019. Durante esse período, na área de $1,35\text{km}^2$ da UFJF, foram observados um total de 108 APs com 36.172 usuários. Usamos um Jupyter Notebook com um kernel `virtualenv Python 3.8.5` de 64 bits e diversos pacotes para explorar visualmente o conjunto de dados. Também utilizamos as bibliotecas `matplotlib` e `seaborn` para gerar os gráficos, enquanto a análise de dados foi realizada usando a biblioteca de análise de dados `pandas`. O conjunto de dados contém 14.295.222 de entradas caracterizadas por quatro atributos:

1. **timestamp**: representa o momento quando um *host* se conecta ao AP específico;

2. **connection_time**: representa o tempo de vida de uma conexão de um *host* no AP específico;
3. **host**: representa a identificação do *host*;
4. **access_point**: representa a identificação do AP.

O conjunto de dados não possui valores faltantes, porém o *connection_time* está no intervalo de $[0, 8.906.006]$ segundos. Embora o valor máximo represente um tempo de conexão de aproximadamente 100 dias, identificamos que este número máximo representa um único *host* sempre ativo, localizado no Restaurante Universitário (RU). Um tempo de conexão de 0 segundos, ou um pouco mais, não faz sentido para qualquer análise de tráfego de rede sem fio. Por isso, o conjunto de dados foi reduzido, removendo todas as conexões com menos de 30 segundos, resultando na nova dimensão de $(7.044.341 \times 4)$, resultando na metade do conjunto de dados original. A escolha de tal quantidade de tempo é que o campus universitário possui uma grande área, e inúmeras instalações, como restaurantes e bibliotecas. Assumimos que um tempo de conexão inferior a 30 segundos significa que um usuário se move pelo campus, ou caminha pelos corredores. Portanto, o usuário pode não gerar tráfego relevante (*i.e.*, transferência de dados), mas é provável que receba notificações (*i.e.*, rede social).

Compreender o padrão de acesso e de usuários no campus é relevante para nosso estudo, e isso é possível por meio de uma análise multivariada. Logo, nossos resultados se concentram em identificar quando, onde e como ocorre a maioria dos acessos. Para iniciar nossa análise, observamos os horários de pico de toda a rede sem fio. A Figura 1 mostra a quantidade de solicitações de acesso que ocorrem ao longo de um dia no campus. O número de conexões é mais expressivo durante o horário comercial, onde a maior parte dos acessos ocorrem das 9h às 18h, e atingem o pico na hora do almoço, entre 11h e 13h. Além disso, a análise mostra que uma parte significativa das solicitações de acesso dos usuários ocorrem de diferentes locais ao longo de um dia. Esse padrão pode ocorrer porque a maioria dos usuários tem aulas em vários locais espalhados pelo campus e possuem uma rotina bem definida em pontos de interesse, como departamentos, bibliotecas e cantinas. De fato, os usuários ficam mais conectados em APs localizados próximos ao caminho das salas de aula e outros poucos pontos de interesse conhecidos. No entanto, os usuários tendem a ficar conectados apenas por curtos períodos. Em média, cada AP nessas rotas conhece cerca de 6k de usuários, e esses usuários permanecem conectados por aproximadamente 5 minutos. Quando um usuário se desconecta, ele pode se reconectar em até 3 minutos e normalmente usa um AP adjacente, caracterizando rotas e padrões de mobilidade e tráfego.

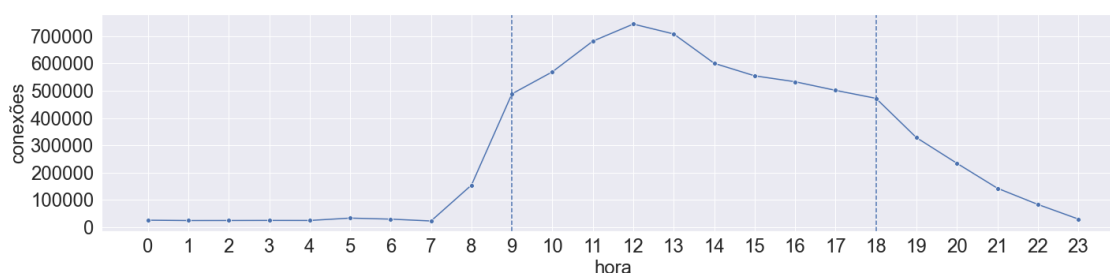


Figura 1. Comportamento de conexões cumulativas ao longo dos dias.

Os APs mais próximos das salas de aula, restaurante universitário e outros pontos

de interesse têm comportamento diferente como ilustrado na Figura 1. Esses APs conhecem uma parcela significativa maior dos usuários, pois esses pontos são *hubs* para alunos, professores e funcionários administrativos, mostrando tempos de conexão mais longos em horários específicos do dia. Por exemplo, a Figura 2 mostra o comportamento de acesso de vários APs. O AP1 e o AP2 estão localizados mais próximos de várias salas de aula, restaurante e cantina, o que significa que esses APs registram a atividade de vários usuários e seus picos de conexão correspondem ao comportamento cumulativo. Por outro lado, AP3, AP4, AP5 e AP6 compõem uma rota entre dois pontos de interesse e têm um comportamento mais estável, com um padrão constante sobre o número de usuários conectados. Esse comportamento ocorre porque esse AP está localizado ao longo de um caminho, e quase ninguém fica nessa localidade.

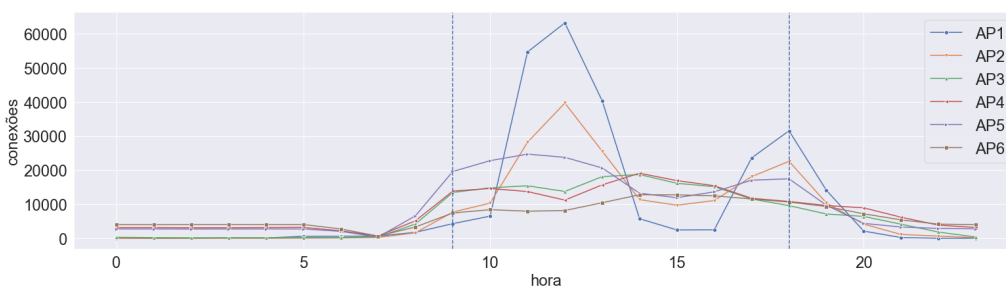


Figura 2. Comportamento do tráfego de rede sem fio ao longo dos dias por ponto de acesso.

Assim, considerando que a parte mais significativa das conexões fica em pontos de interesse, analisamos um conjunto heterogêneo desses APs. A Figura 3 mostra como os APs do Restaurante Universitário compartilham o mesmo padrão, mesmo que a quantidade de conexão seja significativamente diferente. De fato, para construir um modelo capaz de inferir o tráfego da rede sem fio, decidimos focar em APs que atendem ao mesmo comportamento.

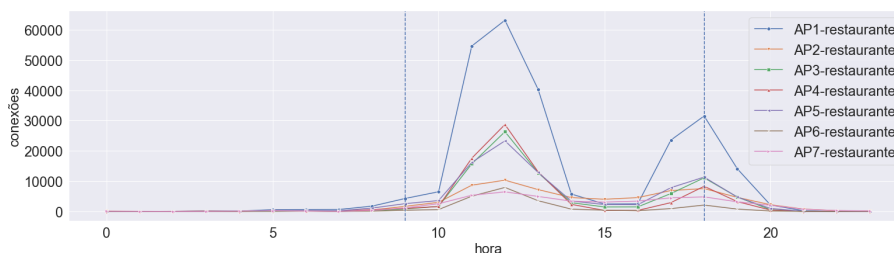


Figura 3. Comportamento do tráfego de rede sem fio ao longo dos dias por ponto de acesso do restaurante.

4. Implementação dos Modelos

Nesta seção apresentamos os modelos utilizados para avaliação no cenário em questão. A implementação deste trabalho pode ser encontrada publicamente no GitHub².

²<https://github.com/LucasFrank/PSO-GA-Comparison>

4.1. PSO

PSO é uma metaheurística baseada em população, tendo seus candidatos a solução chamados de partículas. As partículas se movem no espaço de busca até encontrarem o melhor ajuste. Seus movimentos são guiados por sua própria posição mais conhecida dentro do espaço de busca e a melhor posição conhecida por toda a população. O processo é repetido até que uma solução satisfatória ou critérios de parada sejam alcançados [Kennedy and Eberhart 1995]. As Equações 1 e 2 mostram como a nova posição e velocidade para uma partícula i são calculadas, respectivamente. Na Equação 2, p_{best} é a melhor posição conhecida pela partícula i e g_{best} é a melhor posição visitada por todo o enxame. Também é necessário definir os coeficientes de aceleração c_1 e c_2 , e o coeficiente de inércia w , ambos auxiliando no movimento das partículas. O PSO funciona da seguinte maneira: primeiro, deve-se criar cada partícula, inicializá-la com uma posição aleatória e testar o modelo; e então, um laço principal no algoritmo onde o PSO é executado por N vezes. Em cada rodada, calcula-se uma nova velocidade e posição e testa-se o resultado com o modelo de predição. Para cada teste, o algoritmo verifica se a função objetivo atual da partícula é melhor que anterior. Por fim, verifica-se se essa é a melhor partícula dentre toda a população.

$$s_i^{t+1} = s_i^t + v_i^{t+1} \quad (1)$$

$$v_i^{t+1} = w.v_i^t + c_1.rand.(p_{best} - s_i^t) + c_2.rand.(g_{best} - s_i^t) \quad (2)$$

4.2. Algoritmo Genético

Um AG apresenta uma técnica eficiente de busca e otimização usando seleção natural e reprodução genética inspirada na teoria de Darwin. De acordo com essa teoria, a seleção natural beneficia os indivíduos mais aptos e com maior probabilidade de reprodução. Indivíduos com mais descendentes aumentam a chance de espalhar seu código genético para melhores resultados [Goldberg and Holland 1988]. Seguindo essa ideia, muitos algoritmos computacionais buscam otimizar a solução de um determinado problema por meio de uma população evolutiva com genes artificiais. Para o funcionamento desta técnica, uma população é inicializada com genes aleatórios (no nosso caso, com hiperparâmetros). E então, temos nosso laço principal, que passará por cada indivíduo por um número específico de gerações fazendo as operações de cruzamento e mutação e, finalmente, testando o modelo de predição. O melhor indivíduo é verificado e armazenado ao final desse laço. Apresentaremos o cruzamento e a mutação a seguir.

Cruzamento: Considerando que otimizaremos dois hiperparâmetros, nossos genes são vetores de comprimento dois. A operação de cruzamento é implementada considerando uma probabilidade de 50% pra que cada combinação ocorra. Para o cruzamento ser realizado, o parceiro 1 teria o seu primeiro gene combinado com o segundo gene do parceiro 2, gerando assim uma combinação. Da mesma forma, tem-se a combinação do segundo gene do parceiro 1 com o primeiro gene do parceiro 2.

Mutação: Após cada operação de cruzamento, uma taxa de mutação de 10% pode causar uma operação de mutação nos genes. No nosso caso, um número aleatório é escolhido entre os limites superior e inferior bem definidos para cada hiperparâmetro.

4.3. Perceptron Multicamadas

O MLP é uma Rede Neural Artificial (RNA) na qual todos os neurônios de uma camada estão conectados a todos os outros neurônios das camadas adjacentes. O MLP demonstrou sucesso na predição de cenários cíclicos no passado ([Frank et al. 2019, Ferreira et al. 2019]). Sua arquitetura contém uma camada de entrada, uma ou mais camadas oculta³, e uma camada de saída. A predição é uma resposta unidimensional, ou seja, a predição de tráfego do usuário, a camada de saída contém apenas um neurônio. A Equação 3 descreve o modelo MLP:

$$f(t + 1) = \sum_{i=2}^2 W_i \cdot g(W_{i-1}^T \cdot x_{i-1} + b_{i-1}) + b_i \quad (3)$$

onde: W_i e W_{i-1} são os conjuntos de pesos das camadas i^{th} e $i - 1^{th}$, respectivamente. b_i e b_{i-1} são o viés adicionado às camadas $i + 1^{th}$ e i^{th} , respectivamente. Além disso, $g(x)$ é a função de ativação e, finalmente, x_{i-1} é o conjunto de neurônios da camada $i - 1^{th}$. A arquitetura do modelo MLP é composta por: N sendo o número de neurônios na camada oculta e M sendo o número de neurônios na camada de entrada. Além disso, a taxa de aprendizado é um hiperparâmetro que ajuda a controlar o quanto vai mudar de acordo com o erro do modelo cada vez que o peso do modelo é atualizado. Tanto o número de neurônios quanto a taxa de aprendizado são os hiperparâmetros escolhidos para serem otimizados.

4.4. Árvore de Decisão

A árvore de decisão é um modelo de predição simples e facilmente interpretável. É considerado um dos algoritmos mais comuns usados entre todas as técnicas de ML. Sua arquitetura pode ser definida como nós internos e externos hierarquicamente distribuídos. O nó interno, também conhecido como nó de decisão, é onde ocorre a tomada de decisão por meio de um teste lógico, e assim decide-se qual será o próximo nó filho. Ao contrário do nó interno, o nó externo não possui um nó filho. Portanto, também é conhecido como folha ou nó terminal, tendo um rótulo ou valor associado. [Quinlan 1986]. Quando um conjunto de dados é apresentado a uma árvore de decisão, ele percorre todo o caminho desde o nó raiz (que também é um nó interno) até um nó folha. O resultado é uma classe se for um problema de classificação, ou uma regressão no caso de valor contínuo.

Apesar da árvore de decisão ser um modelo simples, alguns hiperparâmetros precisam ser escolhidos. Como nosso problema é de regressão, o critério escolhido utilizado pelo modelo foi o Erro Quadrado Médio (MSE). Além disso, o hiperparâmetro de profundidade máxima especifica quantos nós existem entre o nó folha mais profundo até o nó raiz. Finalmente, o hiperparâmetro de divisão de amostras mínimas define o número mínimo de amostras para um nó de decisão a ser dividido. Os dois últimos hiperparâmetros foram escolhidos a serem otimizados.

5. Avaliação dos Resultados

Nesta seção apresentamos os resultados obtidos. Primeiramente, introduzimos a metodologia utilizada para avaliação dos modelos, depois os resultados das avaliações dos

³Para a arquitetura do nosso modelo consideramos uma camada oculta.

modelos sobre o *dataset*. Por fim, os resultados aplicando os modelos para a gerência e economia de recursos da rede. Neste último caso, consideramos além do *dataset* real, características reais de tráfego de rede, capacidade dos equipamentos AP e um SLA de 95%.

5.1. Metodologia

Primeiramente, na Seção 5.2 comparamos os modelos usando hiperparâmetros padrão sem o algoritmo de otimização e, em seguida, os modelos propostos (PSO-MLP, GA-MLP, PSO-DT e GA-DT) são comparados entre si. Os modelos foram treinados utilizando da técnica de validação cruzada tendo o número de usuários conectados na região estudada como valor predito. Para essa comparação, as métricas utilizadas serão descritas a seguir:

Coefficiente de Determinação (R^2): A métrica R^2 indica o quanto das variáveis dependentes podem ser explicadas pelo modelo [Glantz and Slinker 2001]. Seu cálculo é definido pela Equação 4, onde y_t é o valor observado, \hat{y}_t é o valor previsto pelo modelo e \bar{y}_t é o valor médio observado, e quando R^2 está mais próximo do valor 1, melhor é o desempenho do modelo.

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} \quad (4)$$

Raiz do Erro Quadrático Médio (RMSE): A métrica RMSE nada mais é do que a raiz quadrada da métrica Erro quadrático médio (MSE), que indica a variação dos erros, também chamada de desvio padrão dos erros [Lehmann and Casella 2006]. Seu cálculo pode ser visto na Equação 5. Seu valor indica um modelo melhor quanto mais próximo de 0 estiver.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (5)$$

Com os modelos já avaliados e o melhor modelo destacado, uma análise em um cenário de gerência de redes é executado para verificar a alocação da quantidade de APs em relação ao número de usuários conectados real/predito. Os experimentos foram realizados pela implementação personalizada na linguagem Python3.

5.2. Avaliação dos Modelos

A Tabela 1 mostra o desempenho de MLP e DT sem que seus hiperparâmetros sejam otimizados. Usando apenas seus hiperparâmetros padrão, os modelos já demonstram bons resultados, tendo a métrica de avaliação R^2 chegando a 93,31% de precisão para o MLP e 89,05% para o modelo DT. O MLP tem apresentado melhor desempenho ao DT, e isso pode ser observado não apenas quando se compara a métrica R^2 , mas também as métricas MAE e RMSE. Ambos apresentaram um erro menor para o modelo baseado em MLP.

Tabela 1. Resumo das métricas para cada modelo sem as metaheurísticas.

	Valor Médio		Desvio Padrão	
	R^2	RMSE	R^2	RMSE
MLP	0.9331	0.0397	0.0044	0.0010
DT	0.8905	0.0508	0.0052	0.0011

Para verificar a eficácia dos algoritmos, as Figuras 4a e 4b, mostram gráficos das métricas de avaliação R^2 e RMSE, respectivamente. Pode-se notar que a curva das

métricas para cada modelo é bastante próxima, e bastante estável ao aumentar o número de iterações. O coeficiente de determinação está mais próximo do valor ideal, apresentando uma acurácia maior que 92% quando se utiliza o modelo baseado em árvore de decisão e maior que 94% quando se utiliza o modelo baseado em MLP. A comparação dos dois algoritmos metaheurísticos (PSO e GA) mostrou a mesma eficiência na otimização dos hiperparâmetros, o que é principalmente verdadeiro para o MLP.

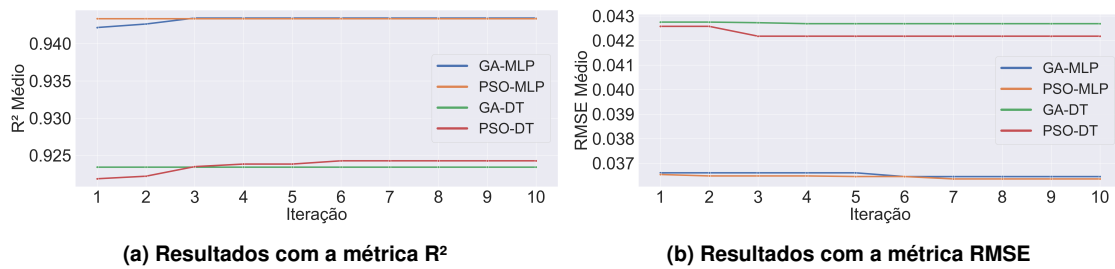


Figura 4. Comparação GA e PSO - otimização de hiperparâmetros MLP e DT.

Os valores do RMSE estão muito próximos de 0, e confirmam o que foi comprovado pela análise de R^2 . De fato, os modelos MLP ainda tiveram um desempenho melhor do que os modelos DT, mesmo que a variação seja de cerca de 0,006%. Os algoritmos metaheurísticos têm desempenho semelhante ao já discutido e, portanto, o resultado da comparação é aproximadamente a mesma. Resumindo, a Tabela 2 é a média das métricas dos quatro modelos após ter sido repetido o processo de treino e teste por 15 vezes. Pode-se observar que as métricas do modelo de predição PSO-MLP são melhores que as métricas dos outros modelos. Ao comparar as Tabelas 2 e 1, podemos ver claramente que ambos os modelos MLP e DT tiveram melhor desempenho quando otimizados pelas metaheurísticas em vez de usar hiperparâmetros padrão, ou ter que ajustá-los manualmente.

Tabela 2. Resumo das métricas para cada modelo utilizando as metaheurísticas.

	Valor Médio		Desvio Padrão	
	R^2	RMSE	R^2	RMSE
PSO-MLP	0.9434	0.0363	0.0014	0.00046
GA-MLP	0.9434	0.0365	0.0015	0.00049
PSO-DT	0.9243	0.0427	0.0014	0.00059
GA-DT	0.9234	0.0422	0.0016	0.00050

5.3. Cenário de Alocação de Recursos para Cobertura em Redes Sem Fio

Nesta seção verificamos a economia de recursos ao mantermos somente APs realmente necessários ligados em uma região específica em um determinado instante. Através do nosso modelo, é possível desativar APs em momentos em que o número de usuários conectados não exige todos os recursos ativos. Porém, é preciso considerar o SLA para manter a QoS e a QoE dos usuários, como veremos adiante.

Considerando um ambiente de rede sem fio atual de larga escala, e a tendência de uso de recursos por aplicações muito populares, como *video streaming* e *cloud-gaming* [Cai et al. 2016], a taxa base de transferência de dados analisada foi definida tal qual em [Carrascosa and Bellalta 2020]. Em Carrascosa e Bellalta, um único usuário

utiliza uma taxa média de dados de $10Mbps$, e considera-se suficiente para utilizar um serviço de mídia de qualidade em $720p$ (*Min*). Por exemplo, tipicamente um AP corporativo Cisco (*e.g.* 1200 Series), pode suportar de 150 a 200 usuários simultaneamente. Mas, vimos que essa quantidade varia em testes reais em um intervalo bem menor, de 20 a 70 usuários [Kafi et al. 2019].

Assim, consideramos um cenário de 1 AP com *link* Ethernet de $1Gbps$ como β , e um limite ($\gamma = 65$) de 65 usuários ($\alpha = [1 - 65]$) – onde cada usuário tem tráfego médio de $10Mbps$ a $15Mbps$ ($\delta = [10 - 15]$). Para ilustrar o cenário real, a região do campus universitário selecionado para o nosso modelo possui cerca de 7 APs. Considerar todos esses APs ativos em momentos de pouco tráfego de rede, ou mesmo sem usuários conectados, leva ao desperdício de recursos. De toda forma, considerando um SLA de $\approx 95\%$ baseado em multiplexação estatística com Distribuição Binomial (p) onde há mais de 10 usuários ativos em 10% do tempo atual, temos a Equação 6.

$$Pr(\alpha > 10) = \sum_{\alpha=11}^{\gamma} \binom{\gamma}{\alpha} \times p^{\alpha} \times (1-p)^{(\gamma-\alpha)} \approx 0.05668 \quad (6)$$

Assim, é possível modelarmos por Programação Linear Inteira (ILP), respeitando a QoE de *Carrascosa e Bellalta* e a Equação 6 para esses usuários ativos, da seguinte forma:

$$\begin{aligned} & \text{maximizar } \sum_{\alpha=20}^{\alpha=70} \alpha \times \delta \\ & \text{sujeito a } \sum_i^{\gamma} \alpha_i \times \delta_i \leq \beta, i = 1, \dots, \gamma \\ & \delta_j \leq \text{Min}, j = 1, \dots, \gamma \end{aligned} \quad (7)$$

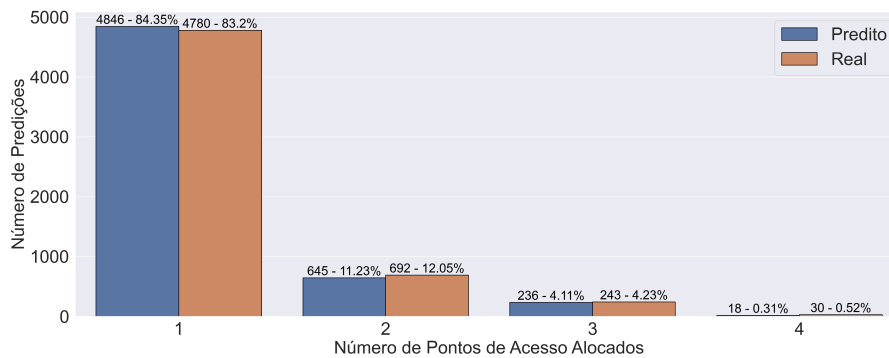


Figura 5. Comparando o valor previsto e o valor real - a quantidade de pontos de acesso necessários.

A Figura 5 mostra o número de APs a serem alocados, comparando o valor previsto e o real de usuários conectados em determinado instante de acordo com o modelo PSO-MLP, com limite $\gamma = 65$ de usuários por AP para um cenário mais restritivo. Os resultados mostram que em $84,35\%$ dos casos, bastaria apenas um AP para atender o número de usuários preditos, e esse está próximo do valor real, que é de $83,2\%$. Ao alocar dois APs, a predição do nosso modelo difere do valor real em apenas $0,82\%$. Além

disso, esse valor diminui ainda mais, quando é necessário alocar três APs, chegando a 0,12%. Finalmente, quando quatro APs são alocados, os valores de predição e real diferem em 0,21%, tendo o número de predições ultrapassando a metade do valor real. Assim, é possível validarmos que, numericamente, nosso modelo atende às expectativas de redução da alocação de recursos mantendo o SLA e o QoE para o usuário final.

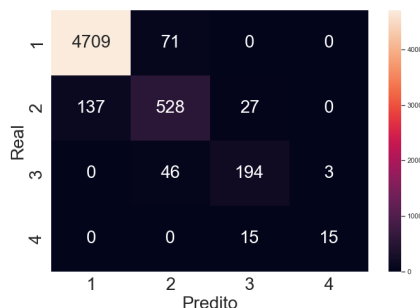


Figura 6. Comparação do valor previsto e verdadeiro da quantidade de pontos de acesso necessários - Matriz de Confusão.

Por fim, a Figura 6 apresenta a matriz de confusão para comparar o número de APs preditos e a serem alocados em relação ao valor real necessário. De acordo com a figura, o modelo acertou 98,51% (4709) das predições ao alocar um AP. Quando são necessários dois APs, a acurácia é de 76,30% (528). Quando se trata de três APs, a acurácia é de 79,83% (194). Já quando são alocados quatro APs, a acurácia chega próximo a 50% (15). Em geral, nosso modelo possui uma acurácia de 94,80% ao prever o número de APs necessários. Adicionalmente, se considerarmos como um problema que comprometa a QoS prever uma quantidade de APs menor que a quantidade real necessária, nesse caso, a acurácia do modelo ficaria próxima a 97%.

6. Conclusões e Trabalhos Futuros

Neste trabalho, investigamos quatro abordagens de predição de tráfego de rede sem fio. A caracterização e os testes foram feitos em uma rede local real. A ideia por trás da análise foi construir 4 modelos a partir de duas metodologias de ML, como o MLP e DT, aplicando o PSO e GA para otimizar os hiperparâmetros. Os algoritmos resultantes foram PSO-MLP, GA-MLP, PSO-DT e GA-DT.

A análise seguiu uma Análise Exploratória de Dados do conjunto de dados, descobrindo um conjunto de APs com o mesmo comportamento de conexão. O conjunto de dados resultante foi avaliado em duas métricas, como o coeficiente de determinação (R^2) e erro quadrático médio (RMSE). Especificamente, o R^2 foi superior a 92% e 94%, respectivamente, usando DT e MLP. Os valores de erro também forneceram suporte a esses resultados. Em todos os casos, a comparação de PSO ou GA como metaheurística mostrou-se com a mesma eficiência na otimização dos hiperparâmetros. Em relação às métricas para cada modelo, o PSO-MLP foi o com melhor desempenho. Além disso, o modelo foi aplicado a um cenário de gerenciamento de recursos de rede, *i.e.*, a alocação de APs de acordo com a predição de tráfego, com acurácia de 94,80% mantendo o SLA de 95%. Outra forma além do melhor gerenciamento dos recursos, a nossa proposta pode oferecer também um guia para possível reposicionamento dos APs com intuito de melhorar o serviço oferecido.

Ainda vemos espaço para estender a pesquisa em várias direções. De um lado, gostaríamos de evoluir os algoritmos, investigar diferentes abordagens e construir um modelo para fazer uma predição de longo prazo do tráfego de rede sem fio. Isso pode ajudar, por exemplo, no melhor planejamento da alocação dos APs. Por outro lado, gostaríamos de investigar a possibilidade da mudança no nível de roteamento dos pacotes. Controladores SDN, como Ryu, ou sistemas operacionais personalizados para APs, como OpenWRT ou DD-WRT, podem auxiliar em testes reais.

Agradecimentos

Este trabalho foi parcialmente financiado pela UFJF, CAPES, FAPEMIG (APQ-00999-18), pelo Projeto Italiano FISIR “La rifunzionalizzazione del Contemporaneo” (CUP: J42F16000600001) e pelo Programa Operacional Nacional Cidades Metropolitanas 2014-2020 (PON METRO) através dos projetos Agorà (CUP F41I17000170006), Cloud@ME (CUP F49E17000040006), ImpleME (CUP F41I17000180006), MeSm@rt (CUP F41I18000230006) e Urbamid+ (CUP F41H18000030006).

Referências

- Bacanin, N., Bezdán, T., Tuba, E., Strumberger, I., and Tuba, M. (2020). Optimizing convolutional neural network hyperparameters by enhanced swarm intelligence metaheuristics. *Algorithms*, 13(3):67.
- Cai, W., Shea, R., Huang, C.-Y., Chen, K.-T., Liu, J., Leung, V. C., and Hsu, C.-H. (2016). A survey on cloud gaming: Future of computer games. *IEEE Access*, 4:7605–7620.
- Carrascosa, M. and Bellalta, B. (2020). Cloud-gaming: Analysis of google stadia traffic. *arXiv preprint arXiv:2009.09786*.
- Casas, P., Seufert, M., Wehner, N., Schwind, A., and Wamser, F. (2018). Enhancing machine learning based qoe prediction by ensemble models. In *IEEE ICDCS*.
- Chih-Lin, I., Sun, Q., Liu, Z., Zhang, S., and Han, S. (2017). The big-data-driven intelligent wireless network: architecture, use cases, solutions, and future trends. *IEEE vehicular technology magazine*, 12(4):20–29.
- Datsika, E., Kartsakli, E., Vardakas, J. S., Antonopoulos, A., Kalfas, G., Maniotis, P., Vagionas, C., Pleros, N., and Verikoukis, C. (2018). Qos-aware resource management for converged fiber wireless 5g fronthaul networks. In *IEEE GLOBECOM*, pages 1–5.
- Deng, T., Wan, M., Shi, K., Zhu, L., Wang, X., and Jiang, X. (2021). Short term prediction of wireless traffic based on tensor decomposition and recurrent neural network. *SN Applied Sciences*, 3(9).
- Ferreira, Y. M., Frank, L. R., Julio, E. P., Ferreira, F. H. C., Dembogurski, B. J., and Silva, E. F. (2019). Applying a multilayer perceptron for traffic flow prediction to empower a smart ecosystem. In *ICCSA*, pages 633–648. Springer.
- Ford, G., Cargan, R., Ahmed, A., Rigney, K., Berry, C., Bucci, D., and Kam, M. (2018). Wireless network traffic disaggregation using bayesian nonparametric techniques. In *CISS*, pages 1–6.
- Frank, L. R., Ferreira, Y. M., Julio, E. P., Ferreira, F. H. C., Dembogurski, B. J., and Silva, E. F. (2019). Multilayer perceptron and particle swarm optimization applied to traffic flow prediction on smart cities. In *ICCSA*, pages 35–47. Springer.

- Fu, Y., Wang, S., Wang, C.-X., Hong, X., and McLaughlin, S. (2018). Artificial intelligence to manage network traffic of 5g wireless networks. *IEEE Network*, 32(6):58–64.
- Glantz, S. and Slinker, B. (2001). *Primer of Applied Regression & Analysis of Variance*, ed. McGraw-Hill, Inc., New York.
- Goldberg, D. E. and Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine learning*.
- Ibrahim, A. A. Z. A., Kliazovich, D., and Bouvry, P. (2016). Service level agreement assurance between cloud services providers and cloud customers. In *IEEE/ACM CC-Grid*.
- Kafi, M. A., Mouradian, A., and Veque, V. (2019). Offline qos association scheme based on clients priorities and demands in wlan networks. In *IEEE GLOBECOM*, pages 1–6.
- Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE.
- Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.
- Li, M., Wang, Y., Wang, Z., and Zheng, H. (2020a). A deep learning method based on an attention mechanism for wireless network traffic prediction. *Ad Hoc Networks*, 107:102258.
- Li, Y., Huang, J., and Chen, H. (2020b). Time series prediction of wireless network traffic flow based on wavelet analysis and BP neural network. *Journal of Physics: Conference Series*, 1533(3):032098.
- Medeiros, D., Cunha Neto, H., Andreoni Lopez, M., Magalhaes, L., Silva, E., Vieira, A., Fernandes, N., and Mattos, D. (2019). Análise de dados em redes sem fio de grande porte: Processamento em fluxo em tempo real, tendências e desafios. *MC-SBRC*, 2019:142–195.
- Prasad, P. S. and Agrawal, P. (2010). Movement prediction in wireless networks using mobility traces.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.
- Saikhu, A., Setyadi, A. T., Purwananto, Y., and Wijaya, A. Y. (2020). Spatio-temporal recurrent neural networks modeling for number of users prediction on wireless traffic networks. In *ICICoS*, pages 1–6.
- Sone, S. P., Lehtomäki, J., Khan, Z., and Umebayashi, K. (2021). Forecasting wireless network traffic and channel utilization using real network/physical layer data. In *EuCNC/6G Summit*, pages 31–36.
- Vieira, A. B., Paraizo, W. N., Chaves, L. J., Correia, L. H., and Silva, E. F. (2021). An sdn-based energy-aware traffic management mechanism. *Annals of Telecommunications*, pages 1–12.
- Yu, Y., Liu, S., Wang, L., Teng, F., and Li, S. (2020). Traffic prediction model based on improved quantum particle swarm algorithm in wireless network. *IOP Conference Series: Materials Science and Engineering*, 768(6):062110.