Previsão da Mobilidade de Usuários em uma Rede Sem Fio Acadêmica

Ana Beatriz Kapps[1], João Victor Balboa[1], Edelberto Franco Silva[1], Gustavo Semaan[2]

[1]Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora (UFJF)
Juiz de Fora – MG – Brasil

[2]Universidade Federal Fluminense (UFF) – Santo Antônio de Pádua – RJ – Brasil

{anabeatrizkapps,jbalboa,edelberto}@ice.ufjf.br, gustavosemaan@id.uff.br

Abstract. *In recent years, the popularity growth of wireless networks has brought computer networking and data science areas even closer. Since many people use Internet-based applications, access to the network has become essential. On the other hand, offering Wi-Fi coverage and availability is not an easy task, requesting planning the network infrastructure. Thus, this work analyzes logs of requests from an institutional Wi-Fi network to improve Internet access and seek to allow an uninterrupted connection during the users' handoff over different access points. In this research, Machine Learning algorithms were applied to predict user displacement, obtaining an accuracy of 76% for groups of users with less mobility.*

Resumo. *O aumento da popularidade de redes sem fio nos últimos anos aproximou ainda mais as áreas de Redes e de Ciência de Dados. Uma vez que muitas pessoas utilizam aplicações ligadas à internet, o acesso a esta se tornou essencial. Por outro lado, disponibilizá-la não é uma tarefa fácil, e deve haver um planejamento em relação a infraestrutura da rede. A fim de melhorar o acesso à internet, buscando permitir uma conexão ininterrupta durante a(s) troca(s) de pontos de acesso, o presente trabalho analisou cerca de dez milhões de requisições realizadas por mais de 36 mil clientes da rede Wi-Fi da UFJF. Foram utilizados algoritmos de Aprendizado de Máquina para prever o deslocamento desses clientes, obtendo-se uma acurácia de 76% para os grupos que possuem menores mobilidades.*

1. Introdução

Uma pesquisa realizada pela Deloitte com 2 mil pessoas, mostrou que a conexão Wi-Fi é o principal meio de conectividade dos Brasileiros, sendo a forma mais utilizada por 84% dos entrevistados [Deloitte 2017]. Isso mostra a importância da manutenção na qualidade deste tipo de conexão. Por isso, o presente trabalho tem como objetivo auxiliar na melhoria do acesso à internet através de conexões Wi-Fi.

Uma das maneiras de melhorar o acesso à internet é permitir conexões sem interrupções, o que é um grande desafio no contexto de redes sem fio, em que os usuários podem se locomover livremente. Com esse deslocamento, o cliente pode se distanciar do ponto de acesso (AP, do inglês *access point*), perdendo intensidade no sinal, até que se perca a conexão. Caso haja outro AP acessível, pode ocorrer uma reconexão automática, de forma que um conjunto de APs seja capaz de cobrir a área desejada.

Uma alternativa para melhorar a conectividade é prever a mobilidade do usuário, ou seja, saber com antecedência para qual AP este mudará. Isso possibilita a realização de uma autenticação pró-ativa, que evita a perda da conexão sem fio, alocando recursos ao usuário mesmo antes de sua chegada. Desta forma, neste trabalho foi realizado um estudo do uso de dados da rede real de Wi-Fi da Universidade Federal de Juiz de Fora (UFJF).

Além da Introdução, a Seção 2 apresenta as etapas de desenvolvimento, o ambiente e os resultados obtidos em relação a predição de mobilidade dos clientes. Por fim, a Seção 3 apresenta as Conclusões e os Trabalhos Futuros.

2. Metodologia

Com o objetivo de buscar e propor melhorias para a conectividade, foram utilizados algoritmos de Aprendizado de Máquina para prever as futuras conexões (que os usuários tendem a realizar). Destaca-se que tais (re)conexões ocorrem de maneira transparente, devido ao uso da rede Wi-Fi durante deslocamentos pelo *campus*.

Para os experimentos computacionais foram utilizados *logs* da rede Wi-Fi da UFJF do período de Junho à Agosto de 2019, que conta com cerca de dez milhões de requisições realizadas por mais de 36 mil clientes distintos em 108 APs. Destaca-se que foi assinado um termo de sigilo para o acesso e a manipulação dos dados, que foram descaracterizados de modo a não ser possível identificar clientes e requisições. Desta forma, a pesquisa pôde seguir respeitando a LGPD (Lei Geral de Proteção de Dados Pessoais).

Com base no volume de dados e os diferentes perfis de deslocamento dos usuários, optou-se por dividi-los em grupos, de acordo com sua mobilidade. Tal organização resultou em cinco grupos (vide Tabela 1), sendo o primeiro de mobilidade fixa, no qual os usuários se conectam em apenas um AP. Este grupo foi desconsiderado nas previsões, uma vez que não houve deslocamento por parte dos clientes.

Tabela 1. Divisão dos clientes em grupos de mobilidade.

Id	Grupo	APs	Requisições	Clientes
1	Fixo	1	44.843	3.830
2	Baixa	2-4	800.535	6.668
3	Média Baixa	5-12	916.203	8.145
4	Média Alta	13-28	2.081.371	8.520
5	Alta	28+	6.187.870	9.006
		Total	10.030.822	36.169

A Figura 1 apresenta o total de clientes (*identidade*) por AP conhecido, em que é possível visualizar a distribuição dos grupos. Com base no gráfico, conforme aumenta a quantidade de APs, diminui o total de clientes. Destaca-se também um considerável quantitativo de clientes que utilizam muitos APs (Grupo *Alta*), com mais de 9 mil clientes que acessam ao menos 28 APs, com picos de até 86 APs para um mesmo usuário.

As tecnologias utilizadas no desenvolvimento dos modelos de previsão se resumem na linguagem *Python versão 3.8.8*, com seus ambientes de desenvolvimento *Spyder 5.1.5* e *Jupyter 6.3.0*, disponibilizados pela plataforma de dados científicos *Anaconda 2.0.3* e na biblioteca *Scikit-Learn* [Buitinck, L. et al. 2013].

O foco deste trabalho foram nas previsões dos grupos de *Baixa* (Id 2) e *Média Baixa* (Id 3) mobilidades. Isso ocorreu considerando a dificuldade de prever a mobilidade

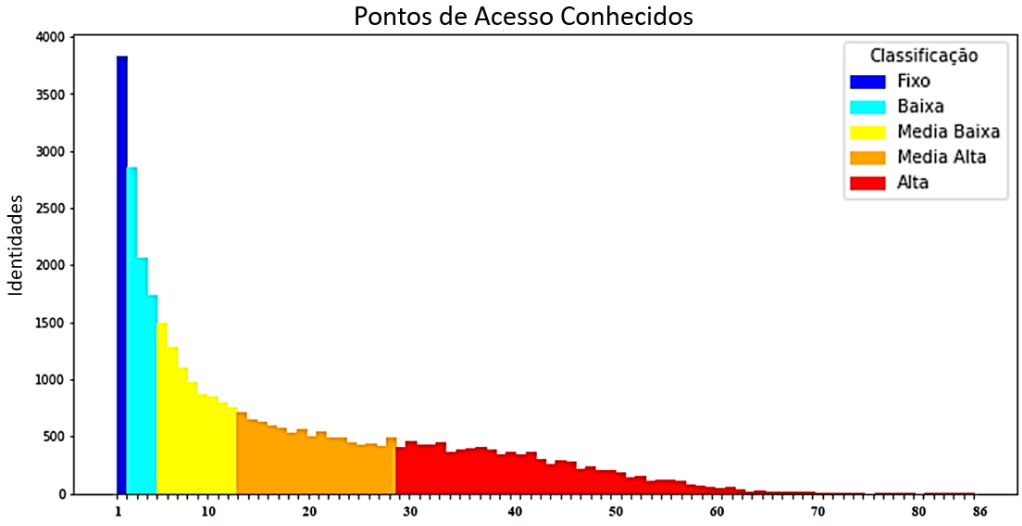

Figura 1. Pontos de acesso conhecidos por usuário de cada grupo.

para os grupos em que a locomoção dos usuários é muito frequente. A Tabela 2 mostra a acurácia dos algoritmos para cada grupo, destacando-se a Árvore de Decisão (*Decision Tree*) e o k-ésimo Vizinho Mais Próximo (*k-Nearest Neighbors*, k-NN), que obtiveram acurácia superior a 70% para o grupo *Baixa* e superior a 60% para o grupo *Média Baixa*.

Tabela 2. Acurácia dos algoritmos.

Algoritmo	Grupo 2	Grupo 3
Decision Tree Classifier	**76%**	**65%**
K Neighbors Classifier	**74%**	60%
MLP Classifier	48%	-
Random Forest Classifier	43%	16%
Naive Bayes (Gaussian)	42%	18%
SGD Classifier	9%	1%

Foram utilizados inicialmente, para cada algoritmo, os parâmetros definidos como padrão pela biblioteca. Em um segundo momento, estes foram alterados empiricamente até chegarem aos melhores resultados, presentes na Tabela 2. Para o classificador baseado na Árvore de Decisão, por exemplo, os parâmetros foram: *criterion* = *entropy, splitter* = *random* e os demais em modo *default*. Já para o k-NN definiu-se a análise de 3 vizinhos ($k = 3$).

3. Conclusão e Trabalhos Futuros

Em experimentos preliminares, antes da organização dos clientes em grupos, a acurácia dos algoritmos era de cerca de 51%. De modo adicional, com o uso de grupos considerou-se como um possível AP qualquer ponto conhecido por um dos clientes desse grupo. Ou seja, ao prever a mobilidade de um cliente, foram considerados APs com os quais este nunca se conectou. Assim, apresenta-se a organização dos grupos em conjunto com o uso de algoritmos de predição como um caminho promissor para a melhoria da qualidade da conectivade.

Portanto, um dos trabalhos futuros consiste na análise sob a ótica do usuário, considerando apenas os APs conhecidos por este. Além deste, pode-se explorar:

- a definição dos grupos de clientes por meio da utilização de técnicas de agrupamento, o que possibilitaria enxergar características não observadas no agrupamento empírico;
- realizar mais experimentos para calibração de parâmetros [Marim et al. 2021];
- a investigação de alternativas que melhorem a previsão para grupos com maior mobilidade, sendo uma delas o aumento da quantidade histórica de APs considerados, ou seja, utilizando n pontos anteriores que o usuário conectou para prever o próximo [Gomes Filho et al. 2021]; e
- a análise por ponto de acesso, que possibilitaria a melhor alocação dos recursos [Frank et al. 2021], removendo, por exemplo, um AP pouco utilizado, ou até mesmo o auxílio na caracterização do comportamento social, obtido através da verificação dos locais e horários com maior concentração de usuários.

Referências

Buitinck, L. et al. (2013). API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122.

Deloitte (2017). Global mobile consumer survey: A mobilidade no dia a dia do brasileiro. Disponível em: "https://www2.deloitte.com/br/pt/footerlinks/pressreleasespage/Global-Mobile-Consumer-Survey.html". Acesso em: 10 out. 2021.

Frank, L. R., de Oliveira, R. M., Vieir, A. B., and Silva, E. F. (2021). Improving a smart environment with wireless network user load prediction. In *26th IEEE Symposium on Computers and Communications (ISCC 2021)*. IEEE.

Gomes Filho, A. R., Cremonezi, B., Nacif, J. A. M., Nogueira, M., Silva, E. F., and Vieira, A. B. (2021). Opportunistic attribute caching: Improving the efficiency of abac in fog-based iot networks. In *ICC 2021-IEEE International Conference on Communications*, pages 1–6. IEEE.

Marim, M. C., Ramos, P. V. B., de Oliveira, R. M., Vieira, A. B., and Silva, E. F. (2021). Caracterização e classificação do tráfego da darknet com modelos baseados em árvores de decisão. In *Anais do XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 127–140. SBC.