

Análise de Associações em Redes Wi-Fi Utilizando Técnicas de Aprendizado de Máquina Multirrótulo para Economia de Energia da Rede

Guilherme H. Apostolo¹, Débora C. Muchaluat-Saade¹, Luiz C. Schara Magalhães²,
Flavia Bernardini¹

^{1,2}Laboratório MídiaCom

¹Instituto de Computação

²Departamento de Engenharia de Telecomunicações

Universidade Federal Fluminense (UFF)

Niterói, RJ – Brasil

{guilherme_apostolo, debora, schara}@midiacom.uff.br, fcbernardini@ic.uff.br

Abstract. *In order to provide wireless access to users, Wi-Fi networks have been extending their capacity, leading to an ever growing power consumption. Power saving in a large-scale Wi-Fi network, without any impact on its user access service, is undoubtedly desired. In this work, we evaluate the performance of Multi-label machine learning methods to provide predictions of access point associations in a real wireless network scenario, using data from SCIFI network of Universidade Federal Fluminense for a period of 6 months. The obtained results showed that the binary relevance method and the decision tree classifier algorithm got the best accuracy of 87.04% for the access point occupancy prediction. The results also showed that a mechanism, as proposed in this work, is capable of saving around 10.34% of all energy consumed by the wireless network.*

Resumo. *Para atender às necessidades de acesso sem fio dos usuários, as redes Wi-Fi têm sido bastante ampliadas e isso leva a custos energéticos cada vez maiores. Economizar energia numa rede Wi-Fi de grande escala, sem nenhum grande impacto no serviço oferecido aos usuários, é sem dúvida algo extremamente desejável. Neste trabalho, analisa-se o desempenho de algoritmos de aprendizado de máquina multirrótulo para fornecer previsões de associações de usuários num cenário real de rede sem fio, utilizando dados da rede SCIFI da Universidade Federal Fluminense (UFF) em um período de 6 meses. Os resultados mostraram que o método relevância binária e o algoritmo árvore de decisão obtiveram a melhor acurácia com um valor de 87,04% para a classificação dos estados de ocupação dos pontos de acesso sem fio. Os resultados também demonstram que um mecanismo como o idealizado por este trabalho é capaz de economizar em torno de 10,34% da energia total utilizada pela rede sem fio.*

1. Introdução

A presença de redes de acesso sem fio em ambientes públicos como shoppings, praças, transportes públicos e universidades é cada vez mais comum [Debele et al. 2014]. Os

usuários desejam estar sempre conectados e consomem, cada vez mais, serviços na Internet [Marsan et al. 2010]. Essa demanda por mais serviços e a presença cada vez maior de usuários na rede levam a um aumento expressivo no consumo de banda da rede, que precisa ser acompanhado por um maior provisionamento de recursos [Ganji et al. 2013].

Para serem capazes de prover conexão a todos os usuários do sistema, com uma banda adequada, as redes sem fio precisam fornecer novos pontos de acesso (APs - *Access Points*) à infraestrutura da rede [Jardosh et al. 2009]. Os novos APs aumentam a capacidade das redes Wi-Fi de suportar múltiplas conexões e de oferecer um melhor serviço de acesso aos seus usuários. Entretanto, a adição de novos pontos de acesso significa um gasto maior com a infraestrutura, não só na sua efetiva instalação, mas também na manutenção dos APs, que inclui, dentre outros custos, o consumo de energia da rede [Lorincz et al. 2010].

Redes sem fio universitárias de grande escala tendem a atender um número enorme de usuários simultaneamente sobre uma área que abriga diversos campi [Magalhães et al. 2013]. Essas redes de grande escala costumam apresentar diversos pontos de acesso com áreas de cobertura sobrepostas em pontos de grande circulação de pessoas durante determinados horários [Fang et al. 2018]. Essa sobreposição permite que a rede absorva um número maior de associações de forma a atender a demanda desses pontos ou locais com densidade elevada de usuários.

Contudo, essas áreas não permanecem extremamente ocupadas durante todo o tempo e algumas outras áreas da rede possuem apenas uma ocupação ocasional [Ganji et al. 2013] [Tang et al. 2012]. No caso de uma rede sem fio de um campus de uma universidade é possível notar que a rede possui um comportamento periódico ao longo das horas e dos dias e que, em muitos momentos, muitos dos APs que compõem a rede ficam em estado de escuta inativo por longos períodos de tempo até que de fato tenham um usuário tentando se associar a eles [Fang et al. 2018].

Segundo [Cui et al. 2013], o consumo de energia associado a uma rede Wi-Fi é considerável. O mecanismo CSMA/CA da rede leva a um consumo ineficiente de energia, uma vez que o consumo de energia em estado de escuta em espera, inativo, é comparável ao do tempo ativamente transmitindo. Com isso é possível dizer, que esses tempos de escuta inativos, em que os pontos de acesso ficam ociosos por longos períodos, significam um desperdício considerável de energia.

Nesse contexto, torna-se interessante a possibilidade de adicionar mecanismos capazes de gerenciar o acionamento e o desligamento de APs em redes sem fio de forma a economizar energia do sistema. Em redes sem fio que possuem uma entidade de controle centralizada capaz de remotamente controlar o acionamento e desligamento das interfaces de comunicação sem fio, pode-se utilizar um modelo capaz de realizar previsões para a presença ou não de associações entre alguma estação e um ponto de acesso da rede, durante uma faixa de tempo, para então executar essas ações e economizar energia sem impactar a capacidade e cobertura da rede. Uma das formas de construir estes modelos e realizar as previsões é utilizando algoritmos de aprendizado de máquina [Fang et al. 2018].

Algoritmos de aprendizado de máquina são capazes de fornecer previsões de ocupação dos diversos APs de uma rede ao longo do dia [Fang et al. 2018]. Neste traba-

lho, analisa-se o desempenho de uma abordagem de aprendizado de máquina multirrotulo de classificação para fornecer previsões de associações, ocupação devido a presença de conexões entre estações móveis e o AP, utilizando dados reais da rede SCIFI da Universidade Federal Fluminense (UFF) em um período de 6 meses.

O modelo de predição treinado com base nos dados reais de ocupação dos APs da rede é capaz de informar ao controlador os horários do dia em que os APs permanecem inativos e assim o controlador é então capaz de desligar a interface sem fio dos pontos de acesso durante estes intervalos de tempo e economizar energia, que seria gasta mantendo a mesma ligada desnecessariamente.

Este cenário foi utilizado como base para realizar os testes de escolha do melhor algoritmo de aprendizado e avaliar o possível potencial de economia de energia utilizando estes algoritmos no cenário proposto. Para possibilitar esta análise, foi criado um *dataset* a partir dos dados coletados dos APs da rede SCIFI de um bloco de salas de aulas da UFF em Niterói durante 6 meses, de abril a setembro de 2018. Foram testados os algoritmos árvore de decisão, KNN e floresta aleatória. Para avaliar o potencial percentual de economia, foi calculada a fração de energia que poderia ter sido economizada [Haratcherev et al. 2009] para o mês de setembro de 2018.

Os resultados obtidos com o *dataset* mostraram que o algoritmo árvore de decisão apresentou o melhor desempenho, com acurácia média de 87%. Contudo, o algoritmo floresta aleatória se saiu melhor, com acurácia média de 81,93%, em uma das técnicas de aprendizado utilizadas nos experimentos, com a técnica cadeia de classificadores. Os cálculos realizados a partir dos resultados obtidos pelo melhor modelo de classificação, para o mês de setembro de 2018, mostram a possibilidade de se economizar uma fração considerável da energia utilizada pela rede SCIFI, em torno de 10% do consumo total da rede do bloco analisado.

O resto deste artigo está estruturado da seguinte forma. Na Seção 2, são discutidos os trabalhos relacionados na literatura. A Seção 3 apresenta as características da rede SCIFI UFF e como os dados foram coletados. A Seção 4 explica detalhes da criação do *dataset* usado para os experimentos assim como suas características. Na Seção 5, é apresentado o detalhamento do cenário utilizado para a avaliação dos modelos. Na Seção 6, são apresentados os experimentos e resultados para a avaliação dos algoritmos e dos métodos de classificação. Por fim, na Seção 7, são apresentadas as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Em [Haratcherev et al. 2009], é desenvolvido um sistema auxiliar de baixa potência que permite o controle da ativação e desativação da interface de rádio de pontos de acesso em redes Wi-Fi. Segundo os autores, a interface de rádio dos pontos de acesso é responsável por uma parcela considerável do consumo energéticos dos mesmos. Os autores então desenvolveram um sistema auxiliar, de baixa potência e fora da banda de funcionamento, que permitia às estações mandarem um sinal responsável por "acordar", ativar a interface de rádio dos pontos de acesso. Os autores, então, estabeleceram métricas de economia de energia, chamada de fator máximo de economia de energia, e realizaram um teste utilizando uma única estação e um ponto de acesso onde a economia de energia máxima medida no ponto de acesso foi de 27,4%. Entretanto, esses sistemas requerem a inclusão

de rádios extras nos pontos de acesso e nas estações móveis, além de softwares adicionais.

O trabalho desenvolvido por [Tang et al. 2012] estuda os efeitos da implementação de um filtro passa faixa e novos métodos de decodificação do sinal de forma a permitir o desenvolvimento de rádios capazes de funcionar com o sistema de acordar e adormecer usando apenas uma interface rádio. Os autores descrevem as diferenças entre técnicas de diminuição do consumo energético para redes de sensores e redes locais sem fio.

Pensando no potencial de economia de energia em redes Wi-Fi empresariais densas, o trabalho de [Ganji et al. 2013] apresenta uma proposta para gerência da topologia da rede de acordo com as necessidades de tráfego presente. A estratégia proposta pelos autores propõe alterações na topologia de acordo com o padrão de utilização apresentado pela rede nos diversos momentos. Essas alterações são o desligamento de alguns pontos de acesso e a alteração da potência por outros de forma a permitir que a rede continue sendo capaz de perceber a presença de estações tentando se associar. Os autores estabelecem que estas redes, usando a estratégia proposta, sejam capazes de gerar uma economia de energia devido ao desligamento total dos pontos de acesso da ordem de 98%.

Os autores em [Jardosh et al. 2009] apresentam o problema do consumo excessivo de energia gerado pelos pontos de acessos em redes Wi-Fi empresariais. Os autores utilizaram os dados obtidos por duas redes prediais distintas, através da coleta de registros do protocolo SNMP durante o período de um mês, e demonstraram que a maior parte dos pontos de acesso da rede permanece inutilizada durante boa parte do tempo e que a maioria desses pontos de acesso ficam ociosos por períodos longos, de 1 hora ou mais. Os autores então introduzem o conceito de recursos sob demanda (ROD). O objetivo principal do ROD é garantir a performance dos usuários da rede enquanto gerencia eficientemente a rede de forma a economizar energia. Os autores então introduzem a política SEAR (*Survey, Evaluate, Adapt, and Repeat*), onde uma controladora central é capaz de realizar os cálculos de demanda e garantia de serviço aos usuários e reconfigurar os pontos acessos da rede de acordo com essas necessidades. O SEAR tenta gerar uma maior economia, maior desligamento de pontos de acesso, de forma a aumentar a eficiência energética da rede.

O trabalho desenvolvido por [Lorincz et al. 2010] desenvolve um mecanismo de gerência da rede capaz de prover uma melhor eficiência energética alterando as potências dos pontos de acesso e desligando/ligando outros pontos de acesso baseado nos indicadores de tráfego da rede.

O trabalho de [Marsan et al. 2010] apresenta uma modelagem estatística para a avaliação da eficácia de políticas ROD para operações de gerência de consumo de energia por uma entidade centralizadora, controladora. Os autores então avaliam a eficácia das técnicas ROD baseadas em atividade e baseadas em tráfego. Para políticas baseadas em atividade foram consideradas duas modalidades, baseada no número de conexões existentes e baseada no número de conexões ativamente transmitindo. Os testes mostraram que as políticas baseadas em atividade se mostraram melhores que as baseadas em tráfego e que a política baseada em conexões existentes se mostra a opção mais atraente devido ao seu bom desempenho e simplicidade.

Visando realizar um controle eficiente do consumo energético de uma rede sem

fio de larga escala, os autores em [Fang et al. 2018] apresentam a proposta do mecanismo ACE (*AP Control with Energy saving*). O ACE é capaz de dinamicamente controlar o status de funcionamento, ligado ou desligado, de cada um dos APs, garantindo a cobertura da rede e gerando uma economia de energia na ordem de 70%. Para basear sua decisão de ligar ou desligar os APs, a quantidade de usuários associados em cada APs é predita utilizando o algoritmo de aprendizado de máquina único rótulo floresta aleatória. Cada AP possui seu próprio modelo de previsão, que é treinado e avaliado utilizando um histórico de dois meses de dados obtidos através do armazenamento de pacotes SNMP. Os modelos levam em consideração os atributos dia da semana, feriado ou não e hora do dia como entrada para fornecer sua previsão de ocupação dos APs, ocupado ou inativo, e atingem, em média, uma acurácia de 90%.

A partir dessa previsão, o ACE determina que os APs que possuem um tempo de inatividade superior a T períodos, de tempo de meia hora, devem ser desligados. Os autores avaliam como diversos valores de T afetam a economia de energia, porcentagem de energia economizada pelo desligamento dos APs sobre o total consumido, e a área de cobertura, razão de clientes que conseguiram se conectar aos APs que se encontravam ligados sobre o total de clientes que tentaram se conectar aos APs da rede durante os período de análise.

Diferentemente das propostas apresentadas por [Jardosh et al. 2009] [Lorincz et al. 2010] e [Ganji et al. 2013], o presente trabalho avalia a utilização dos modelos de aprendizado de máquina multirrótulo para fornecer a previsão de ocupação dos APs da rede. A utilização dos modelos de aprendizado de máquina permitem uma maior facilidade e generalização na forma como a previsão ocorre, podendo expandir o horizonte de previsão dos estados para o dia inteiro baseado num histórico de estados passados. Isso retira a necessidade da realização mais constante de medições pela rede e da realização cálculos frequentes para todos os APs, o que poderia se mostrar computacionalmente muito custoso para uma rede de grande escala.

A análise feita por este trabalho avalia o desempenho, acurácia, dos algoritmos de aprendizado utilizando os métodos relevância binária e o método de classificadores em cadeia. Isso torna este estudo diferente do apresentado por [Fang et al. 2018] quanto às técnicas de aprendizado de máquina e métodos de treinamento que podem ser utilizadas. Este estudo também desenvolveu e utilizou um *dataset* próprio baseado em dados reais de acesso sem fio, para treino e teste dos classificadores, sendo mais longo do que o usado no estudo de [Fang et al. 2018], porém que conta com um número menor de pontos de acesso sendo registrados.

3. Rede SCIFI UFF e Coleta de Dados

A rede sem fio SCIFI UFF é composta por um controlador inteligente, também chamado de controlador SCIFI, e APs de baixo-custo, SOHO, operando sobre o *firmware open source* OpenWRT. O controlador SCIFI (Sistema de Controle Inteligente para redes sem Fio) é o elemento central de coordenação e monitoramento de todos os elementos da rede sem fio e usa o eduroam como forma de autenticação. O controlador coordena as funções de coleta de informações, seleção de canais e potência de transmissão dos APs [Magalhães et al. 2013].

O controlador SCIFI é extensível e permite customizações adicionais, porém já

permite a instalação de um grande número de APs na rede e provê algoritmos para controle de potência de transmissão e seleção de canal dos APs para maximizar a utilização espectral. O sistema hoje em dia também já permite o desligamento das interfaces de comunicação sem fio dos APs, função essa essencial para a futura implantação do mecanismo de economia de energia utilizado no cenário proposto deste trabalho. A rede SCIFI permite uma redução expressiva no custo da rede, o que facilita a instalação de redes muito maiores com o mesmo orçamento.

O projeto da rede SCIFI foi desenvolvido pela Universidade Federal Fluminense (UFF) com financiamento da Rede Nacional de Pesquisa (RNP). O SCIFI foi desenvolvido com o intuito de permitir a instalação, operação e gerência de uma rede sem fio de baixo custo e de *software* livre [Magalhães et al. 2013].

A rede SCIFI está presente em diversos campi da UFF em Niterói, com diversos APs espalhados pelos prédios. Foram selecionados, para este estudo, os APs do bloco H do campus Praia Vermelha da UFF em Niterói. O bloco H é um prédio composto unicamente de salas de aulas. Por esse motivo, a sua ocupação ocorre apenas decorrente da utilização das salas para aulas ou aplicação de exames, diferente de outros prédios que possuem espaços destinados a outros fins como sala de professores, coordenações de curso e laboratórios, por exemplo. A rede SCIFI possui 28 APs no bloco H distribuídos pelos 5 andares do prédio.

Os dados utilizados neste estudo foram obtidos a partir dos registros de eventos dos APs coletados e armazenados pelo controlador SCIFI. Cada AP envia um arquivo de texto contendo as informações de eventos de controle e gerência de camada física e de enlace registradas por eles. Os dados são então requisitados e armazenados semanalmente pelo controlador, durante a madrugada. Foram utilizados os dados coletados durante 6 meses, de abril a setembro de 2018.

Os arquivos contêm diversas informações, porém, neste estudo, como o interesse é apenas prever estados de inatividade ou atividade dos APs, foram filtrados os registros para conterem apenas as informações de associação, desautenticação e desassociação. Os eventos de associação e desassociação marcam o início e o término da transmissão de dados entre uma estação móvel e um AP. Dessa maneira, pode-se classificar que qualquer AP que possua pelo menos uma estação móvel associada durante um intervalo de tempo x estará em atividade durante este intervalo de tempo x , onde o início do estado de atividade ocorreu no momento em que a estação móvel se associou ao AP e o término desse estado ocorreu no momento em que a estação se desassociou.

Percebeu-se, porém, que nem sempre o evento de desassociação era registrado pelos APs. Entretanto, a etapa de desautenticação sempre ocorria em par a etapa de autenticação. Além disso, sempre que os eventos de desassociação e desautenticação ocorriam em par, eram registrados para a mesma estação, ambos ocorriam em intervalos de tempo muito próximos, da ordem de 1 segundo. Por isso, utilizou-se o registro de desautenticação como marco do término da conexão, entre a estação móvel e o AP, na falta do registro de desassociação da estação móvel ao AP. Uma análise mais profunda seria necessária para verificar o porquê da falta dos registros das desassociações, mas uma das possíveis causas é a movimentação do usuário da estação móvel para fora da área de cobertura da rede.

4. Construção do *Dataset* e suas Características

A partir da filtragem dos registros para conter apenas as informações relativas à conexão das estações móveis aos APs, do bloco H, foi construído um *dataset* que permitiu a organização dessas informações e a construção de um histórico diário de ocupação dos APs durante intervalos de tempo.

Baseando-se nos trabalhos de [Fang et al. 2018] e [Sangogboye et al. 2016], organizou-se o *dataset* mostrado na Tabela 1, mostrando um histórico de utilização dos diferentes APs do bloco H no período de 6 meses, de abril a setembro de 2018. Cada dia foi dividido em intervalos de tempo de 10 minutos, o que totaliza 144 intervalos de tempo por dia. Vale salientar que seria possível a escolha de outros valores para o parâmetro de divisão do dia e criação dos intervalos de tempo, porém optamos por seguir o parâmetro de 10 minutos já utilizado no trabalho de [Sangogboye et al. 2016].

Tabela 1. Parte do *dataset* construído, mostrando os atributos e histórico de ocupação para os diversos APs

Ano	Mês	Dia	Apid	Feriado	Dia da Semana	Time0	Time1	...	Time143
2018	Apr	8	21	F	Sunday	0	0	...	0
2018	Apr	8	164	F	Sunday	0	0	...	0
2018	Apr	8	223	F	Sunday	0	0	...	0
.									
.
.									
2018	Jul	18	269	T	Wednesday	0	0	...	1
2018	Jul	18	276	T	Wednesday	0	0	...	0
2018	Jul	18	277	T	Wednesday	0	0	...	1

Cada instância do *dataset* representa os estados de ocupação de um determinado AP do bloco H para um dia específico. Os atributos ano, mês e dia indicam a data específica em que os estados de ocupação de um AP estão sendo apresentados. O atributo Apid indica a qual dos APs do bloco H o histórico de ocupação pertence. O atributo dia da semana indica o respectivo dia da semana. O atributo feriado/férias é booleano e indica se o dia representado, na instância, é um dia letivo, com aulas (F - *false*), ou não, um dia de férias ou feriado (T - *true*).

Cada atributo Time é booleano e indica a inatividade, valor 0, ou atividade, valor 1, do AP no intervalo de tempo indicado. Para um AP ser classificado como inativo é preciso que nenhuma estação tenha se conectado a ele, mesmo que apenas por um segundo, durante todo o intervalo, caso contrário o AP terá seu estado para o intervalo de tempo considerado como ativo, mesmo que tenha apresentado apenas um conexão por um período de tempo muito inferior à janela de análise do intervalo. Cada atributo Time representa um intervalo de tempo específico, onde atributo Time0 representa o intervalo de tempo de 00:00 às 00:10 e os demais seguem a ordem crescente, sempre acrescentando mais 10 minutos com relação ao atributo anterior na sua janela de tempo.

A Figura 1 mostra alguns comportamentos interessantes da rede SCIFI no bloco H a partir da análise do *dataset*. Nessa figura, é possível observar as médias diárias e semanais, de ocorrência do estado de ocupação dos APs durante os intervalos de tempo de

10 minutos. Os valores foram normalizados pelo número total de ocorrências do estado para um determinado intervalo de tempo, de forma a permitir uma melhor visualização e escala do gráfico. Sendo assim, os valores para a ocorrência do estado de ocupação são valores reais entre 0 e 1, sendo que quanto mais próximo de 1 for o valor de ocorrência do estado, maior é o número de ocorrências observadas no *dataset* para aquele estado no intervalo de tempo analisado.

A Figura 1 permite observar que a média varia de acordo com o dia da semana, sendo mais difícil a ocorrência de momentos de inatividade durante dias da semana do que nos dias de final de semana e a média de ocupação costuma ser maior nas terças, quartas e quintas. Outro comportamento capturado na Figura 1 é a diferença de ocupação média entre os sábados e domingos. Isso se deve à eventual utilização do prédio aos sábados para a aplicação de exames e outras atividades especiais.

Com base na Figura 2, é possível observar que existe sempre um grande volume de APs inativos durante a madrugada. A ocupação média, em que os APs encontraram-se no estado ativo durante os horários de 00:00 a 6:00 é muito pequena. Também é possível observar uma maior média de ocupação quando comparamos os dias úteis com feriados ou férias escolares ou dias do final de semana. A ocupação dos APs é muito maior em dias úteis do que em feriados, dias durante as férias escolares e finais de semana. Entretanto, a Figura 2 também mostra que a ocupação média para finais de semana é menor do que a apresentada em feriados ou férias escolares. Mesmo que de início este comportamento pareça estranho, este pode ser explicado pela utilização das salas de aulas durante o período de férias para cursos de férias aos alunos, por isso existe uma ocupação mais constante durante as férias do prédio por esses cursos do que a ocupação experimentada aos finais de semana.

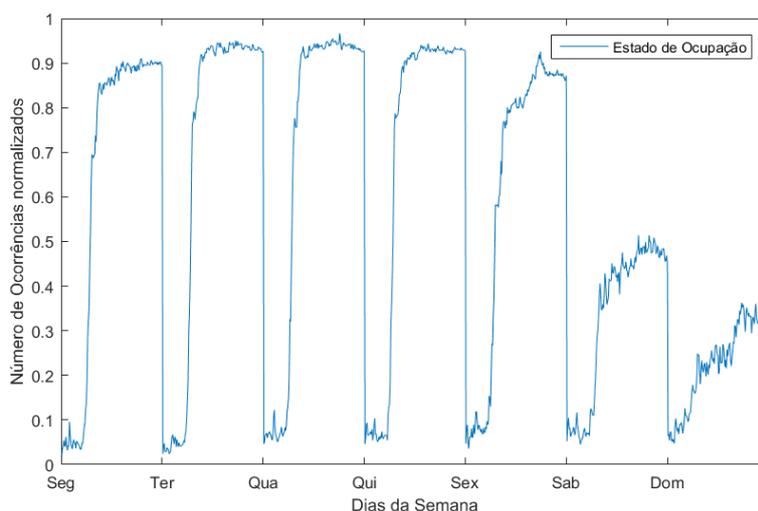


Figura 1. Média do número de ocorrências de estado de ocupação dos APs na rede SCIFI do bloco H durante os dias da semana

Como o bloco H é um prédio de salas de aula, esses comportamentos eram esperados uma vez que a ocupação se dá majoritariamente devido à presença ou não de aulas no prédio. Uma vez que não há aulas nos dias de férias, feriados ou finais de semana, e nem nos horários entre 22:00 e 7:00, a maioria dos APs fica inutilizado durante esses

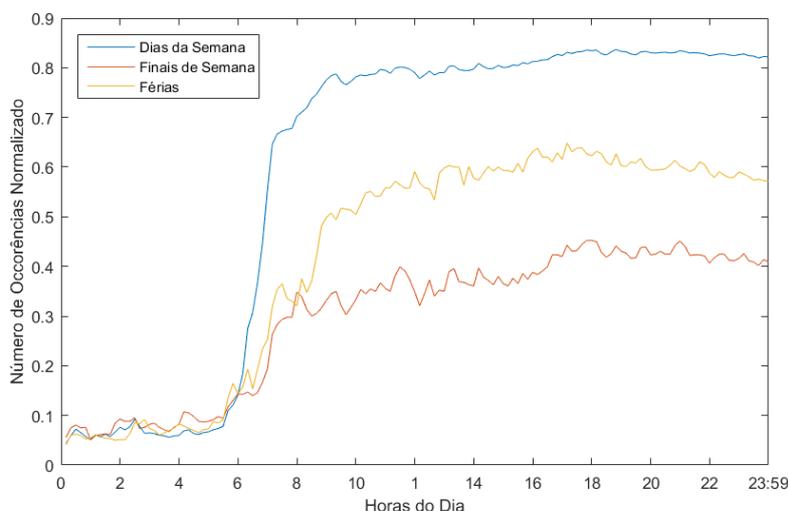


Figura 2. Média do número de ocorrências de estado de ocupação dos APs na rede SCIFI do bloco H durante os intervalos de tempo de um dia

horários. Mesmo assim existem aparelhos, como computadores, e funcionários, como segurança e limpeza, presentes no prédio durante esses dias, de forma que não tem como simplesmente assumir que não é possível ter algum dispositivo conectado nestes dias e horários.

5. Cenário Proposto para a Avaliação dos Algoritmos de Aprendizado de Máquina

A partir do *dataset* e das avaliações apresentadas na Seção 4, é possível identificar o potencial desperdício de energia gerado pelo funcionamento dos APs em momentos de inatividade. Dessa forma, um mecanismo capaz de identificar esses períodos de inatividade na rede SCIFI e, então, desligar as interfaces de comunicação sem fio dos APs que permanecem inutilizados durante um intervalo de tempo, poderia gerar uma economia de energia na rede, com o mínimo de impacto aos seus usuários.

Essa estratégia de utilizar os recursos de rede, os APs nesse caso, sob demanda através do seu acionamento e desligamento de acordo com a sua utilização pode reduzir significativamente o consumo de energia [Debele et al. 2014]. Uma das melhores estratégias de decisão sobre a utilização de recursos sob demanda é baseando-se no número de conexões existentes [Marsan et al. 2010]. Nesta estratégia a decisão sobre a utilização do recurso fica atrelada ao número de conexões estabelecidas entre as estações móveis e o AP. Essa tomada de decisão pode ocorrer baseada na estimativa, previsão, do número de conexões ativas que o AP terá em momentos futuros utilizando algoritmos de aprendizado de máquina [Fang et al. 2018].

Utilizando algoritmos de aprendizado de máquina, pretende-se obter as previsões para os futuros estados de ocupação, existência ou não de pelo menos uma estação móvel conectada, dos APs durante os intervalos de tempo, de 10 minutos. Cada modelo seria, então, capaz de prever os estados de ocupação de um AP para todos os seus 144 intervalos de tempo do dia.

Serão utilizadas duas técnicas de aprendizado de máquina multirrótulo para a construção do classificador, a relevância binária (BR - *Binary Relevance*) e a técnica de classificadores em cadeia (CC). Ambas técnicas apresentadas para lidar com o problema são consideradas transformações do problema multirrótulo em um problema único rótulo. Essa característica permite a utilização de algoritmos único rótulo.

Na técnica da relevância binária, constrói-se um modelo para cada intervalo de tempo sem levar em consideração os demais intervalos de tempo. Esta abordagem é mais simples e não explora a possível correlação existente entre os rótulos. Já no método de classificadores em cadeia, esta correlação dos rótulos é explorada, uma vez que o modelo do próximo intervalo de tempo levará em consideração os valores obtidos para os intervalos de tempo anteriores.

Cada AP terá um classificador próprio, que será composto pelos modelos de previsão para cada um dos 144 intervalos de tempo, e será capaz de prever os estados de ocupação, ativo e inativo, para um dia inteiro. Os experimentos utilizam os algoritmos de classificação KNN, árvore de Decisão e floresta aleatória devido aos bons resultados já apresentados em outro trabalho relacionado [Fang et al. 2018]. Os atributos de entrada utilizados serão os campos Apid, feriado e dia da semana do *dataset*. Os meses do *dataset* de abril a agosto serão usados para treino e o mês de setembro será usado para teste.

A métrica de avaliação utilizada é a acurácia. A acurácia permite avaliar a capacidade do modelo de fornecer uma previsão correta. É analisada a média da acurácia fornecida pelos modelos dos APs para cada um dos 144 intervalos de tempo.

É possível avaliar a utilização das previsões de ocupação fornecidas pelos algoritmos de aprendizado de máquina, num cenário onde estes resultados são usados por um mecanismo de economia de energia capaz de gerenciar o acionamento e desligamento das interfaces de comunicação sem fio baseado na informação de ocupação. Caso o modelo preveja que um AP fique inativo pelo próximo intervalo, este deve então ficar com sua interface desligada até que seja previsto a ocupação do mesmo novamente. Baseados nos dados disponíveis e neste mesmo cenário, é possível comparar os resultados obtidos com os algoritmos de aprendizado de máquina contra uma política básica, onde os APs são mantidos desligados durante a madrugada, entre 0 e 6am, e ligados durante o resto do dia, 7am em diante. Esta comparação contra uma política simples mostra o ganho que os algoritmos de aprendizado de máquina têm sobre um algoritmo de gerenciamento baseado em limiares de operação e sua capacidade de prever mudanças de comportamento em momentos do dia.

A avaliação feita para este cenário permite, também, estimar a possível economia de energia capaz de ser obtida pela rede SCIFI do bloco H da UFF durante o período de um mês caso um mecanismo como o idealizado fosse de fato implementado. Vale a pena ressaltar que o mecanismo de economia de energia idealizado é simples e teve como propósito fornecer uma base de avaliação para calcular um percentual de economia de energia e dar uma solução simples que pode ser mais facilmente implementada futuramente na rede SCIFI, que já possui os mecanismos necessários para o desligamento das interfaces sem fio, o que acelera e muito o desenvolvimento prático e a implementação real do mecanismo proposto nesse cenário de rede. Uma desvantagem clara do mecanismo idealizado para avaliação neste cenário frente aos outros, na lite-

ratura [Jardosh et al. 2009][Lorincz et al. 2010], é a perda de área de cobertura da rede durante os períodos de desligamento das interfaces de comunicação sem fio.

6. Experimentos e Resultados

A Figura 3 mostra os resultados obtidos para a média da acurácia dos APs. A acurácia dos APs foi obtida pelos classificadores para os intervalos de tempo do dia utilizando a API scikit-learn [Buitinck et al. 2013]. Pela Figura 3, é possível perceber que o método relevância binária (BR) obteve uma média da acurácia dos APs melhor que o método de classificadores em cadeia (CC). Também é possível perceber que há uma queda brusca na acurácia dos classificadores para os intervalos de tempo entre 35 e 50. Esse horário representa o horário de abertura do prédio e das primeiras aulas da manhã.

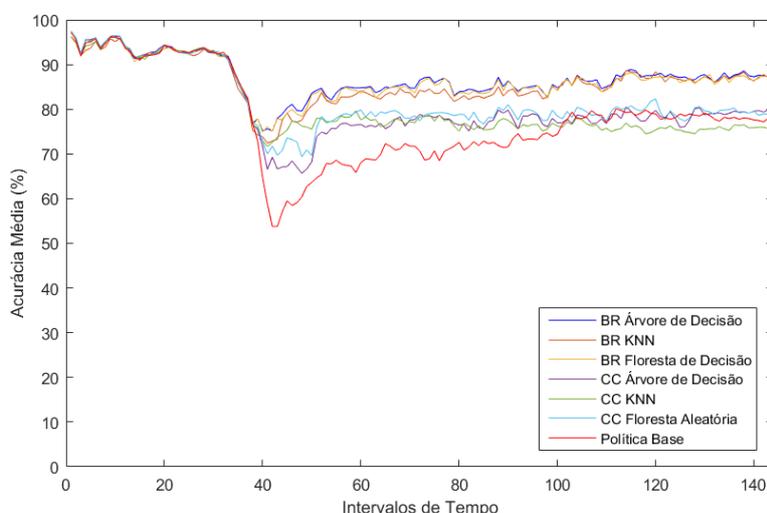


Figura 3. Média da acurácia dos APS dos diferentes algoritmos e métodos obtida para rede SCIFI do bloco H durante os intervalos de tempo de um dia

O desempenho pior dos classificadores em cadeia, possivelmente, ocorre devido à mudança repentina nos estados dos APs nesse momento da manhã. Essas mudanças causam um aumento na dificuldade dos classificadores em prever corretamente os valores, o que leva a uma queda na acurácia dos APs. Essa perda da acurácia introduz uma maior possibilidade de erro e consequentemente afeta todo o restante da cadeia. Uma vez que, no método relevância binária, as informações obtidas pelos classificadores de atributos alvo anteriores, ou seja, em momentos anteriores, não são levadas em conta, então essa propagação de possíveis erros não acontece.

A Figura 3 mostra que todos os algoritmos na abordagem BR apresentam melhor acurácia que a política base para o período de 7am em diante. Para o período entre 7 e 10am, o algoritmo de aprendizado tem acurácia em média 20% maior. Entre 10 e 23:59 esta diferença é de 10%. Os testes mostraram que para o período de 0 a 6:59, a acurácia entre ambos é similar. Os resultados também mostram que os algoritmos árvore de decisão e floresta aleatória conseguiram a melhor acurácia média.

A Tabela 2 mostra os resultados das acurácias médias obtidas para os algoritmos de aprendizado e os métodos utilizados. A acurácia média representa a acurácia geral,

acurácia levando em conta todos os valores preditos para o conjunto de teste, do algoritmo para o método. O método BR utilizando o algoritmo árvore de decisão conseguiu a melhor acurácia média geral dos experimentos, que ficou em torno de 87,04%.

Tabela 2. Média das acurácias dos APs obtidas pelos diferentes algoritmos de aprendizado e métodos de classificação

Algoritmos	Acurácia Média BR	Acurácia Média CC
Árvore de Decisão	87,04%	80,02%
KNN	86,00%	80,52
Floresta Aleatória	86,67%	81,93%

O trabalho de [Haratcherev et al. 2009] apresenta uma formulação matemática, apresentada na Equação 1, para o cálculo do percentual de economia de energia gerado pelo desligamento da interface de comunicação sem fio de um AP durante um intervalo de tempo. Através dessa fórmula e dos resultados obtidos pelo classificador, é possível fornecer o percentual de energia possível de ser economizado no cenário de teste proposto para o mês de setembro de 2018.

Na Equação 1 os termos P_{ext_on} e P_{ext_off} representam o valor da potência em Watts medida, na fonte de alimentação externa ao AP, para os casos da interface de comunicação sem fio ligada e desligada, respectivamente. Os termos t_{on} e t_{total} representam o tempo total em que os APs permaneceram funcionando com a interface de comunicação sem fio ligada e o período de tempo total da avaliação dos APs, respectivamente. Ao calcular estes valores, atinge-se um percentual, que representa a fração de energia que foi economizada pela rede devido ao desligamento das interfaces de comunicação durante o período de tempo avaliado.

$$F = \frac{P_{ext_on} - P_{ext_off}}{P_{ext_on}} \left(1 - \frac{t_{on}}{t_{total}}\right) \quad (1)$$

Com os resultados do classificador utilizando o método BR e o algoritmo árvore de decisão, melhor resultado obtido, foi possível prever que os APs do bloco H permaneceriam no estado inativo durante 43,20% do tempo para todo o mês de Setembro. É possível também estimar que os APs do bloco H permaneceriam no estado inativo durante 28,47% do tempo para todo o mês de Setembro caso a política base fosse utilizada. As medições realizadas para este experimento, também permitiram calcular quais seriam as potências de consumo de um AP do mesmo modelo utilizado em todo o bloco H. As medições demonstram que este modelo possui uma variação de potência entre o modo de operação com a interface de comunicação ligada e desligada em torno de 0,30W.

Os valores das potências de consumo com a interface ligada e com a interface desligada para o modelo de AP utilizado na rede estão presentes na Tabela 3. Estão presentes também, a fração máxima de energia possível de ser economizada, F_{max} , que representa a economia que seria gerada caso a rede permaneça com as interfaces de comunicação desligada durante todo o tempo e o percentual de energia capaz de ter sido economizado pela rede SCIFI UFF do bloco H para o mês de setembro de 2018 utilizando o classificador com os melhores resultados, FML_{Set} e utilizando a política base, FTR_{Set} .

Tabela 3. Valores de potência consumida pelo modelo de AP utilizado, percentual máximo de economia de energia e percentual de economia de energia obtido pelo experimento para o mês de Setembro de 2018 e utilizando a política base

P_{ext_on} (W)	P_{ext_off} (W)	F_{max} (%)	FML_{Set} (%)	FTR_{Set} (%)
1,111	0,845	23,93	10,34	6,81

A Tabela 3 nos mostra que seria possível atingir um percentual de economia de energia de 10,34% para o mês de setembro, no cenário apresentado. Isso significa que caso o mecanismo proposto seja utilizado na prática, este seria capaz de economizar uma razão considerável de energia de uma rede sem fio. A Tabela 3 também mostra que a utilização que um classificador, obtido através de técnicas de aprendizado multirrotulo, apresentaria um percentual de economia de energia maior do que o obtido através de um mecanismo baseado em limiares de horários como o da política base. Estes resultados confirmam que os classificadores não são só mais corretos em suas previsões como são capazes de fornecer um maior percentual de economia de energia.

7. Considerações Finais

Através da criação de um *dataset* a partir dos registros de eventos dos APs armazenados no controlador SCIFI, foi possível notar o comportamento periódico de ocupação da rede e sua modificação ao longo do dia, da semana e de períodos letivos e férias. Os resultados demonstraram um perfil de ocupação dos pontos de acesso condizente com os períodos de realização de aulas e atividades no prédio. Foi possível perceber, no entanto, que devido a algumas características esporádicas, como realização de concursos e provas em sala, e de alguns equipamentos específicos, como os computadores e outros dispositivos de rede do prédio que permanecem ativos durante grandes períodos de tempo, há alguns comportamentos não periódicos que tornam o perfil de ocupação distinto em alguns momentos.

A avaliação dos resultados mostra que o método relevância binária e o algoritmo árvore de decisão obtiveram a melhor acurácia para a previsão de ocupação dos APs, com um valor de 87,04%. Vale também notar que um mecanismo, como o idealizado por este trabalho, teria sido capaz de economizar em torno de 10,34% da energia total utilizada pela rede durante o mês de setembro de 2018.

Isso demonstra que o futuro desenvolvimento de um mecanismo baseado na análise apresentada seria capaz de prever com considerável acurácia os estados de ocupação dos APs da rede e gerar economia de energia ao sistema. Isso possibilitaria também a gerência da topologia da rede de acordo com as necessidades de conexões presentes. O mecanismo iria realizar alterações na topologia de acordo com o padrão de utilização apresentado pela rede nos diversos momentos. Vale ressaltar porém que a aplicação deste mecanismo em outros cenários de redes, como em shoppings e praças, necessitaria de sua avaliação com dados obtidos nestes cenários. Esta avaliação seria uma interessante extensão do trabalho apresentado.

Outro aspecto interessante a ser analisado é a extensão da base de dados do *dataset* para incluir um período maior de tempo a ser utilizado para as etapas de treino e teste. Outra extensão do trabalho é incluir também um estudo sobre a capacidade dos algoritmos de prever a quantidade de conexões ativas durante um intervalo de tempo.

Referências

- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., and Varoquaux, G. (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.
- Cui, Y., Ma, X., Wang, H., Stojmenovic, I., and Liu, J. (2013). A survey of energy efficient wireless transmission and modeling in mobile cloud computing. *Mobile Networks and Applications*, 18(1):148–155.
- Debele, F. G., Li, N., Meo, M., Ricca, M., and Zhang, Y. (2014). Experimenting resource-demand strategies for green w lans. *ACM SIGMETRICS Performance Evaluation Review*, 42(3):61–66.
- Fang, L., Xue, G., Lyu, F., Sheng, H., Zou, F., and Li, M. (2018). Intelligent large-scale ap control with remarkable energy saving in campus wifi system. In *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 69–76. IEEE.
- Ganji, F., Budzisz, Ł., and Wolisz, A. (2013). Assessment of the power saving potential in dense enterprise w lans. In *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 2835–2840. IEEE.
- Haratcherev, I., Fiorito, M., and Balageas, C. (2009). Low-power sleep mode and out-of-band wake-up for indoor access points. In *2009 IEEE Globecom Workshops*, pages 1–6. IEEE.
- Jardosh, A. P., Papagiannaki, K., Belding, E. M., Almeroth, K. C., Iannaccone, G., and Vinnakota, B. (2009). Green w lans: on-demand wlan infrastructures. *Mobile Networks and Applications*, 14(6):798–814.
- Lorincz, J., Capone, A., and Bogarelli, M. (2010). Energy savings in wireless access networks through optimized network management. In *IEEE 5th International Symposium on Wireless Pervasive Computing 2010*, pages 449–454. IEEE.
- Magalhães, L. C. S., Balbi, H. D., Corrêa, C., Valle, R. D. T. d., and Stanton, M. (2013). Scifi—a software-based controller for efficient wireless networks.
- Marsan, M. A., Chiaraviglio, L., Ciullo, D., and Meo, M. (2010). A simple analytical model for the energy-efficient activation of access points in dense w lans. In *Proceedings of the 1st international conference on energy-efficient computing and networking*, pages 159–168. ACM.
- Sangogboye, F. C., Imamovic, K., and Kjærgaard, M. B. (2016). Improving occupancy presence prediction via multi-label classification. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6. IEEE.
- Tang, S., Yomo, H., Kondo, Y., and Obana, S. (2012). Wake-up receiver for radio-on-demand wireless lans. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):42.