

Uma Infraestrutura para o Monitoramento e Predição de Rotas e Paradas de Ônibus no Transporte Universitário

Paulo Miranda e Silva Sousa¹, José Robertty de Freitas Costa¹
Abilio Castro e Silva¹, Wilton Ribeiro Crispim Neto¹
Carla Ilane Moreira Bezerra^{1,2}, Emanuel Ferreira Coutinho^{1,2}

¹Campus de Quixadá – Universidade Federal do Ceará (UFC)
63.902-580 – Quixadá – CE – Brasil

²Programa de Pós-Graduação em Computação (PCOMP)

{paulomirandamss, abilio, robertty, wiltonribeiro}@alu.ufc.br

{carlailane, emanuel.coutinho}@ufc.br

Abstract. *Due to the growth of urbanization, cities have faced social, economic and environmental transformations. In addition, many vehicles currently have several sensors and actuators, capable of performing not only the sensing of the condition of vehicles, but also the environment around them, and this data can be used for various services. The environment of a large university may resemble urban environments, considering that these institutions compare to cities in various aspects, especially in relation to infrastructure problems. The objective of this work is to develop an infrastructure for the monitoring and prediction of bus routes and stops in university transportation. Tests were performed on real bus routes, and the results, even if preliminary, were attractive and have potential for research.*

Resumo. *Devido ao crescimento da urbanização, cidades têm enfrentado transformações sociais, econômicas e ambientais. Além disso, muitos veículos atualmente possuem diversos sensores e atuadores, com capacidade de executar não só o sensoriamento da condição dos veículos, mas também do ambiente ao seu redor, sendo que esses dados podem ser utilizados para diversos serviços. O ambiente de uma universidade de grande porte pode se assemelhar com ambientes urbanos, considerando que essas instituições se comparam às cidades em vários aspectos, principalmente em relação aos problemas de infraestrutura. O objetivo deste trabalho é desenvolver uma infraestrutura para o monitoramento e predição de rotas e paradas de ônibus no transporte universitário. Foram realizados testes em rotas de ônibus reais, e os resultados, mesmo que preliminares, se mostraram atrativos e com potencial para pesquisa.*

1. Introdução

A partir do crescimento populacional e da urbanização, cidades de maneira geral têm enfrentado transformações sociais, econômicas e ambientais, tais como o aumento da desigualdade social, acesso limitado aos serviços públicos básicos, problemas de mobilidade urbana e segurança [de M. Neves et al. 2017]. Neste contexto, o conceito de Cidades Inteligentes (*Smart Cities*) surge como um novo paradigma que visa encontrar

soluções sustentáveis para esses problemas crescentes. Diversas soluções foram propostas e desenvolvidas para os campi universitários, considerando que essas instituições se comparam às cidades em vários aspectos, principalmente em relação aos problemas de gestão, falta de infraestrutura e consequente insatisfação dos seus usuários [Sungur et al. 2015, Saad et al. 2018, Feng et al. 2018].

Atualmente, muitos veículos possuem uma gama muito grande de sensores e atuadores, com capacidade de executar não só o sensoriamento da condição dos veículos, mas também do ambiente ao seu redor [Rodrigues et al. 2018]. As informações coletadas por esses sensores podem ser utilizadas para diversos serviços, que poderão levar um maior conforto e segurança aos passageiros e aos próprios condutores do veículo.

A cobertura de toda uma área urbana com uma rede de sensores pode ser proibitivamente custosa [Cruz et al. 2018]. Uma opção para a redução dos custos e ainda assim obter uma significativa cobertura espacial é utilizar nós móveis de sensoriamento, onde cada nó é responsável por sensoriar uma região. Uma forma mais barata de prover mobilidade nos nós de sensoriamento é embarcá-los em ônibus urbanos [Castro et al. 2017].

Em um cenário ainda mais específico, onde alguns campi universitários possuem convênios com empresas ou prefeituras para auxílio em algum tipo de serviço, como transporte de alunos, uma solução que otimizasse rotas e paradas, gerenciasse horários e lotação dos ônibus, ou aumentasse ou diminuísse a frota conforme a necessidade, seria de grande valor para a comunidade acadêmica [Feng et al. 2018].

Nesse contexto, o objetivo deste trabalho é desenvolver uma infraestrutura para o monitoramento e predição de rotas e paradas de ônibus no transporte universitário. A partir dessa infraestrutura, informações sobre o transporte dos universitários de duas universidades, Universidade Federal do Ceará (UFC) e Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), ambas localizadas na cidade de Quixadá, Ceará, serão tratadas em uma aplicação e sistema embarcado. Também será possível apresentar por meio da infraestrutura desenvolvida a localização em tempo real dos ônibus universitários e prever o horário do próximo ônibus em uma determinada parada. Foram utilizadas técnicas de aprendizado de máquina para prever o horário do próximo ônibus em um determinado ponto de parada, obtendo-se uma margem de erro muito pequena a partir dos dados coletados nas rotas.

O artigo está dividido nas seguintes seções, descritas a seguir. Na Seção 2, são apresentados conceitos aplicados a este trabalho. Na Seção 3, são apresentados alguns trabalhos relacionados presentes na literatura. Em seguida, na Seção 4, é detalhada a solução proposta para o sistema de monitoramento e predição do transporte universitário. Na Seção 5, são apresentados o planejamento do experimento, a execução dos testes, os resultados e análises do experimento. Por fim, a Seção 6 apresenta as conclusões deste trabalho.

2. Fundamentação Teórica

Nesta seção, apresentamos os principais conceitos para um melhor entendimento do presente trabalho.

2.1. Microcontrolador ATmega328p

Os microcontroladores são pequenos circuitos integrados que possuem memória, processador, periféricos de entrada e saída e entre outros componentes. Os sistemas microcontrolados estão cada dia mais presentes em diversos equipamentos do nosso cotidiano. Segundo [Martins 2005], os microcontroladores estão presentes em quase tudo o que envolve a eletrônica, diminuindo o tamanho, facilitando a manutenção e gerenciando tarefas internas de aparelhos eletroeletrônicos. A redução de custos é um dos principais fatores que contribuem para a popularização e utilização dos microcontroladores.

Neste trabalho, utilizamos o microcontrolador ATmega328p¹ pelo fato de possuir diversas plataformas de desenvolvimento e facilidade de gravação. A plataforma Arduino foi escolhida para o desenvolvimento da aplicação. A plataforma, apresenta um ambiente de desenvolvimento de algoritmos próprio e *upload* através da porta USB de qualquer computador, tornando prático o desenvolvimento de aplicações.

2.2. Aprendizado de Máquina

Aprendizado de Máquina é uma área da computação que tenta prever situações futuras tendo como base informações anteriores. Segundo [Monard and Baranauskas 2003], o objetivo da área de Aprendizado de Máquina é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática.

É notório que a área de Aprendizado de Máquina está em constante crescimento e também são encontradas diversas aplicações para a mesma. Esta pesquisa irá trabalhar com aprendizado supervisionado, em específico, com algoritmos de regressão para tentar prever o horário do próximo ônibus em uma determinada parada.

Os algoritmos de aprendizado de máquina supervisionado testados foram *Linear Regression*, *Decision Tree*, *Random Forest* e o *XGBoost*. O modelo de predição foi feito na linguagem de programação Python, onde os três primeiros algoritmos são da biblioteca do *scikit-learn* e o último algoritmo da biblioteca *xgboost*. Esses algoritmos tem se mostrado bastante eficientes em diversos problemas de aprendizado. Muitos desses algoritmos já são utilizados em diversos trabalhos da literatura que lidaram com problemas semelhantes ao que se propõe esse trabalho. Um exemplo disso é o trabalho [Yamaguchi et al. 2018] que tenta prever o tempo de viagem de ônibus em diversas rotas distintas. Para a avaliação do modelo de aprendizado final foram utilizados as métricas MAE (*Mean Absolute Error*) e RMSE (*Root Mean Square Error*), afim de determinar qual melhor algoritmo a ser usado.

3. Trabalhos Relacionados

Existem na literatura diversas soluções para sistemas IoT de monitoramento do transporte urbano, como por exemplo, os sistemas de monitoramento de ônibus. No trabalho de [Castro et al. 2017] são apresentados diversos trabalhos da literatura para sistemas de monitoramento de ônibus, onde são destacadas as tecnologias usadas, implementação, dificuldades e problemas nesses sistemas. Dentre as tecnologias utilizadas pelos trabalhos

¹http://microchip.com/downloads/en/DeviceDoc/Atmel-7810-Automotive-Microcontrollers-ATmega328P_Datasheet.pdf

identificados destaca-se o GPS (*Global Positioning System*) e o GSM (*Global System for Mobile Communication*). Estes trabalhos utilizam o GPS como a tecnologia de rastreamento em conjunto com GSM como tecnologia de comunicação. A utilização conjunta entre GPS e GSM também será utilizada neste trabalho.

Dentre estas aplicações de monitoramento urbano, têm se destacado os sistemas de monitoramento de ônibus dos campi universitários, pelo fato dessas instituições se compararem às cidades em vários aspectos, principalmente em relação aos problemas de gestão, falta de infraestrutura e consequente insatisfação dos seus usuários [Sungur et al. 2015, Miranda et al. 2017, Saad et al. 2018, Feng et al. 2018].

[Sungur et al. 2015] apresentam um sistema inteligente de informações sobre passageiros com paradas para permitir que administradores monitorem o sistema de transporte público e possibilite pessoas que utilizam este sistema observem simultaneamente as informações sobre a localização e o estado desses veículos. Como estudo de caso, utilizou-se as paradas em um campus universitário e outros pontos urbanos da cidade. As tecnologias utilizadas neste trabalho foram um dispositivo embarcado ARM Micro2440 e a posição do GPS, onde os dados são coletados e mantidos em um banco de dados.

No trabalho [Miranda et al. 2017] é apresentado a construção de um sistema de localização de ônibus com registro automático de rotas e acesso em tempo real através de um aplicativo móvel com custo acessível, onde são utilizados os conceitos de sistemas embarcados, redes e geolocalização para aplicar o paradigma de IoT como uma forma de ajudar e aumentar o uso de transporte público em geral. A solução utilizada no trabalho foi *Amazon AWS* para *back-end* em nuvem, unificando os dados transmitidos pelo GPS do ônibus para aplicativos Android. O terminal instalado no ônibus consiste em um Raspberry PiTM usando um receptor GPS e o módulo HSDPA para transmissão de dados. Os dados transmitidos são recebidos no aplicativo Android e são plotados em tempo real usando Google MapsTM, permitindo ao usuário saber onde está o seu ônibus desejado, e até mesmo saber o seu ponto em sua rota. A solução proposta em [Miranda et al. 2017] se assemelha com a proposta deste trabalho, pois existem algumas diferenças entre módulos utilizados no hardware que compõe o sistema embarcado mas a funcionalidade de tais módulos em relação à solução aqui apresentada é a mesma.

No trabalho [Saad et al. 2018] é apresentado um sistema de monitoramento em tempo real dos ônibus dos campi. Semelhante a este trabalho, é utilizado o GPS na infraestrutura proposta para capturar dados como latitude, longitude e velocidade. Esse sistema, assim como muitos da literatura, tenta prever o tempo de chegada nas paradas dos ônibus baseado em somente distância e velocidade do ônibus. Como um diferencial da infraestrutura proposta neste trabalho, são utilizados algoritmos de aprendizado de máquina para previsão.

O trabalho de [Feng et al. 2018] realizam um estudo sobre a aplicação da localização de nós da IoT e a conscientização de localização dos ônibus de um Campus Universitário para melhoria da previsão da chegada dos ônibus e integra tecnologias RFID e Zigbee de redes sem fio em um sistema de barramento inteligente. O sistema pode localizar e monitorar o estado de execução do barramento em tempo real, fornecendo para que o barramento colete as informações do tempo nas estações, e melhoria da qualidade, eficiência e satisfação do serviço de ônibus.

Sistemas de transportes inteligentes possibilitam a utilização de sensores e GPS no monitoramento dos sistemas de transportes públicos em cidades inteligentes [Neves et al. 2018]. A captura e processamento desses dados permite, a princípio, a utilização do transporte público com confiabilidade e previsibilidade, o que melhoraria a qualidade de vida da população urbana. Entretanto, diversos fatores provocam que esses dados sejam insuficientes ou de baixa qualidade para uso em tempo real. O trabalho analisa dados obtidos por meio de colaboração coletiva (*crowdsourcing*) para complementar informações. Além disso, o trabalho propõe um modelo de análise de confiabilidade dos dados coletados para o sistema de transporte público (por ônibus) do município de São Paulo.

A dinâmica espacial e social de uma cidade sofre constantes mudanças ao longo do tempo [Wen et al. 2018]. Em consequência, a malha viária e o sistema de transporte público necessitam ser otimizados continuamente para atender a diversas demandas. Uma alternativa para reduzir custos e impactos na avaliação de soluções é o emprego de simuladores com modelos consistentes com a realidade. Este trabalho processou dados de deslocamento e de planejamento do sistema de ônibus de São Paulo para melhorar o modelo de movimentação de ônibus usado pelo InterSCSimulator, um simulador escalável para cidades inteligentes. Um modelo de movimentação baseado em dados reais do serviço de ônibus de São Paulo foi apresentado para tornar o simulador mais eficaz ao recriar cenários de mobilidade urbana.

As condições climáticas adversas representam um fator emblemático afetando negativamente a qualidade do transporte público, especialmente em regiões sujeitas a chuvas [Vianna et al. 2018]. Este trabalho apresenta relações entre eventos de chuva e o comportamento da velocidade média dos ônibus de transporte público na cidade do Recife. Para isto, utilizou-se estatística descritiva na análise dos dados do itinerário, posicionamento e velocidade dos ônibus em relação aos dados de precipitação em estações pluviométricas. Análises de locais conhecidos por problemas de trânsito em dias de chuva foram detalhadas no artigo.

4. Infraestrutura para Monitoramento e Predição de Rotas e Paradas de Ônibus no Transporte Universitário

A solução proposta para a infraestrutura de monitoramento e predição de ônibus no transporte universitário foi dividida em três fases: (i) prototipação do sistema embarcado; (ii) obtenção dos dados; e, (iii) construção do produto final.

De maneira geral, na primeira fase do projeto construiu-se um protótipo do sistema embarcado visando testar os componentes do sistema e o código desenvolvido. Com o protótipo em funcionamento, iniciou-se a obtenção dos dados da rota. Por fim, em posse dos dados, um modelo de aprendizado de máquina supervisionado foi projetado e desenvolvido, e o produto final foi construído.

Na Figura 1, pode-se visualizar a arquitetura do sistema e como cada parte é interligada. No elemento Sistema Embarcado é apresentado o dispositivo responsável por emitir os dados utilizados pelo sistema. Esse dispositivo possui um módulo GPS responsável por capturar dados como coordenadas (latitude e longitude), horário e velocidade. Outro módulo desse sistema é o GSM/GPRS. Tal módulo é responsável por enviar os dados capturados pelo GPS via requisição POST para um servidor WEB. Para enviar

os dados, utiliza-se um chip nano SIM que permite fazer o envio dos dados via internet da rede telefônica. Esses módulos são controlados pelo microcontrolador *Atmega 328p* da *Microchip Technology*.

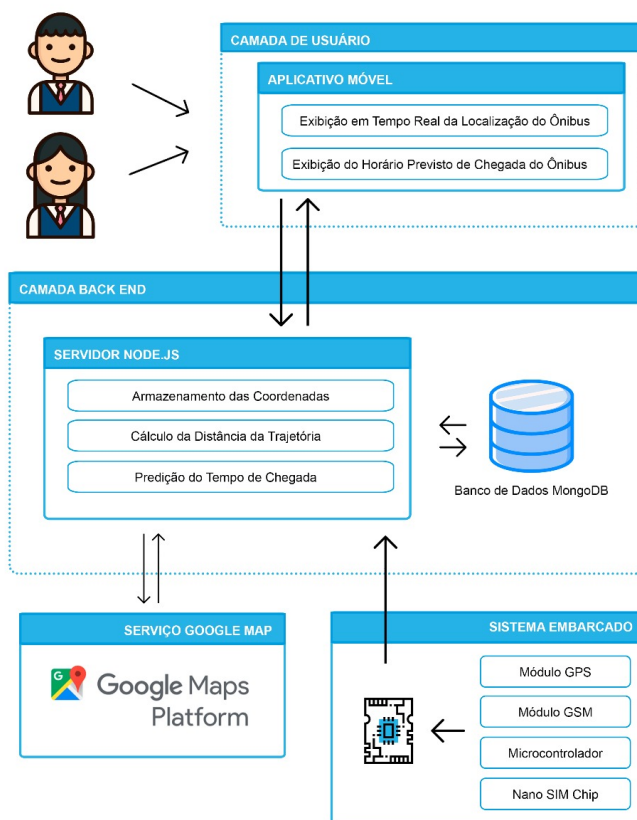


Figura 1. Arquitetura do Sistema

Outra camada importante do sistema é o *Back End*. Essa camada é responsável por receber todas as requisições POST do sistema embarcado e armazená-las em um banco de dados Mongo. Além disso, é nessa camada que é efetuada a predição do tempo de chegada do próximo ônibus em uma determinada parada e todo o suporte necessário para a aplicação do usuário.

Comunicando-se diretamente com o *Back End*, tem-se a camada de usuário. Tal camada ainda encontra-se em fase de desenvolvimento. Esta possuirá uma aplicação *mobile* que exibirá em tempo real a localização dos ônibus e o horário de chegada do próximo ônibus. Além disso, nessa aplicação, as informações sobre o transporte universitário serão concentradas em um *feed* do aplicativo.

Por fim, a camada *Google Maps Service* dará suporte tanto ao *Back End* quanto ao usuário. No *Back End*, serão extraídas informações da cidade como, por exemplo, a menor distância entre dois pontos que respeita todas as sinalizações das ruas. Na aplicação do usuário será utilizada a API do *Google Maps* para que os usuários do aplicativo tenham uma visão mais amigável da posição atual do ônibus, podendo acompanhar em tempo real o deslocamento do mesmo.

Na Figura 2 são apresentados os componentes do sistema embarcado e como eles estão interligados. No rótulo 1 tem-se a antena utilizada pelo módulo GSM/GPRS. No rótulo 2 tem-se um regulador de tensão que converte 6V-12V para 3.3V e 5V. Em 3, é visualizado o módulo GPS *VK2828U7G5LF TTL* que é responsável pelos dados de coordenada, hora e velocidade. Em 4 é possível identificar o módulo GSM/GPRS *SIM 800L v.2* que é responsável por enviar uma requisição POST para um servidor com as informações do módulo GPS. No rótulo 5 tem-se o microcontrolador Atmega 328p responsável por gerenciar todo o sistema. Por fim, em 6, tem-se 3 LEDs (Diodo Emisor de Luz) que são utilizados para exibir o estado atual do sistema. LED vermelho ligado significa que o sistema está sem conexão de internet. LED amarelo aceso sozinho, significa que o dispositivo está aguardando dados do GPS. LED amarelo e verde ligado, os dados estão tentando ser enviado pelo GSM/GPRS. LED verde aceso sozinho, os dados foram enviados com sucesso para o servidor.

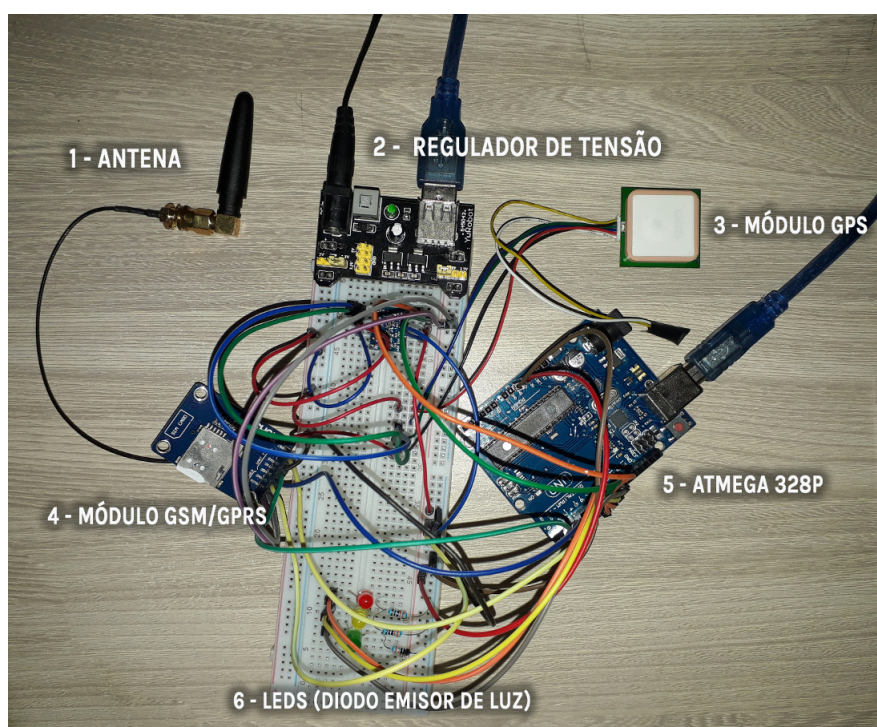


Figura 2. Arquitetura do Sistema Embarcado

5. Planejamento e Execução do Experimento

Nesta seção o planejamento do experimento é descrito, sua execução é detalhada e os resultados apresentados e discutidos.

5.1. Planejamento do Experimento

Os experimentos têm como objetivo verificar a precisão das coordenadas do módulo GPS e o comportamento do sistema em áreas de sinal telefônico fraco.

As coletas de dados foram realizadas em dois trajetos: (i) percurso feito pelo ônibus no trajeto Praça José de Barros - UFC/IFCE; e (ii) percurso feito pelo ônibus no

trajeto UFC/IFCE - Praça José de Barros. Apesar de serem pontos de início e fim invertidos para as duas rotas, os trajetos são diferentes. O sistema embarcado foi transportado durante os trajetos.

Os dados coletados pelo sistema foram as coordenadas em latitude e longitude, horário de emissão e a velocidade do veículo. Todos esses dados são capturados pelo módulo GPS descrito na seção anterior.

Com esses dados, alguns algoritmos da literatura irão prever o horário de chegada do ônibus na parada. Os algoritmos aplicados foram: *Decision Tree*, *XGBoost*, *Linear Regression* e *Random Forest*. Foram utilizadas as métricas MAE (*Mean Absolute Error*) e RMSE (*Root Mean Square Error*) para análise dos resultados.

Para visualização e análise dos dados, mapas com os trajetos, mapas com os pontos de coleta, gráficos de barras e tabelas foram desenvolvidos para suportar o experimento.

5.2. Execução do Experimento

A coleta de dados foi realizada no período de 11/03/2019 até 15/03/2019. No total, 12 coletas foram realizadas, 6 no percurso Praça José de Barros - UFC/IFCE (Figura 3) e 6 na rota UFC/IFCE - Praça José de Barros (Figura 4). Os mesmos ônibus usados no transporte universitário foram utilizados no experimento. O sistema foi ligado antes de entrar no ônibus e desligado logo após a saída. Assim, tentou-se manter os dados o mais próximo possível da realidade.

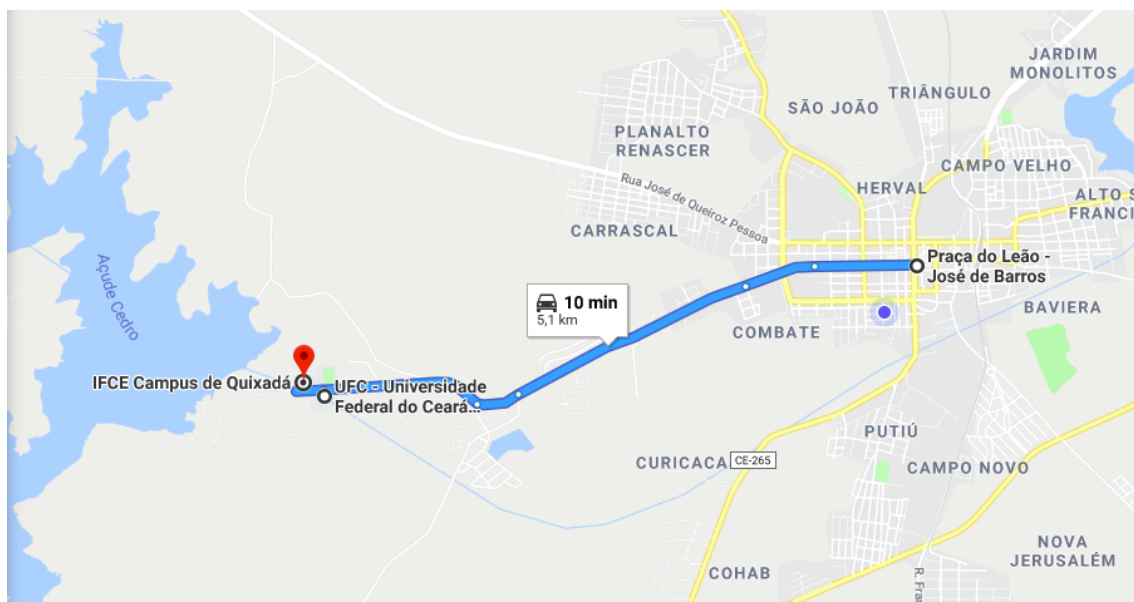


Figura 3. Rota no sentido Praça - UFC/IFCE. Fonte: Google Maps

Durante o experimento a coleta foi parcialmente afetada por falha de conexão do módulo GSM/GPRS à rede. Mesmo com essa condição adversa, a coleta de dados foi proveitosa. Após a coleta de dados, foi montada a base de dados afim de construir o modelo de aprendizado de máquina. Para isso, os dados foram separados em 70% para treino e 30% para teste. Como dito previamente, os algoritmos de aprendizado supervisionado testados foram *Linear Regression*, *Decision Tree*, *Random Forest* e o *XGBoost*.

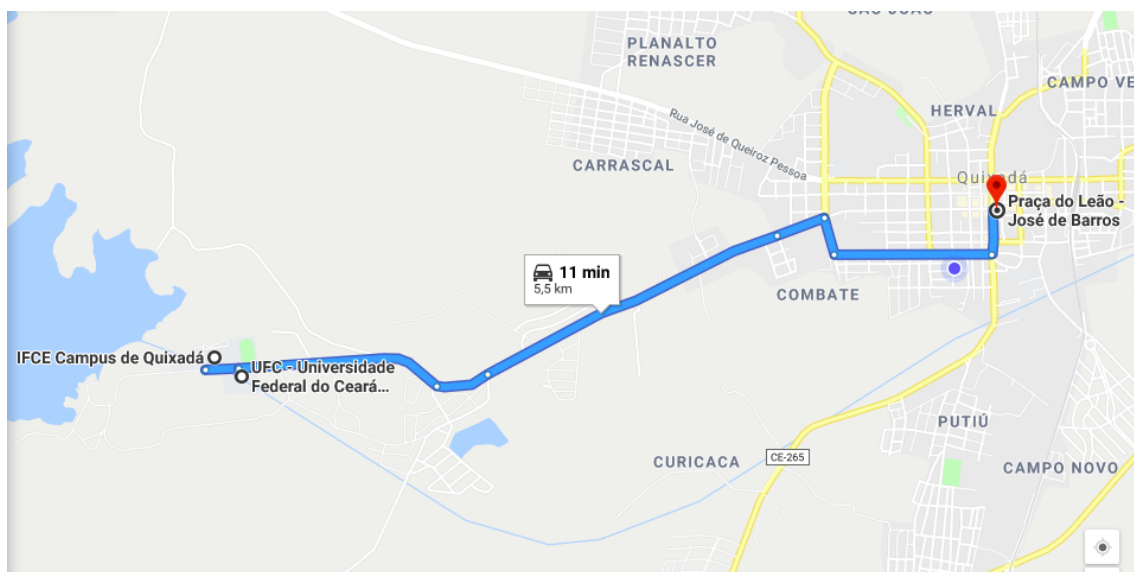


Figura 4. Rota no sentido UFC/IFCE - Praça. Fonte: Google Maps

Para avaliar o desempenho dos algoritmos foram utilizadas duas métricas: MAE (*Mean Absolute Error*), descrita na Equação 1 e RMSE (*Root Mean Square Error*), descrita na Equação 2. Essas técnicas são amplamente utilizadas para testar a eficiência de algoritmos de regressão.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y - \hat{y}| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y - \hat{y})^2} \quad (2)$$

As Equações 1 e 2 descrevem, respectivamente, as métricas MAE e RMSE, onde y é o valor real e \hat{y} é o valor predito pelo algoritmo de aprendizado de máquina.

O módulo GPS consegue captar informações como latitude, longitude, horário e velocidade. Foram fixados os pontos de latitude e longitude das paradas do ônibus e calculados o tempo que demorou para o ônibus chegar a uma determinada parada. Foi inserida a distância euclidiana entre o ponto de partida e o ponto de chegada (ponto que representa uma parada do ônibus). Por fim, foram retiradas as coordenadas de latitude e longitude (tanto de partida como destino), pois as mesmas não se mostraram serem atributos relevantes quando os seus valores brutos são inseridos para treinar o algoritmo. Por fim, obteve-se três atributos que são eles: a velocidade do ônibus (em km/h), o horário (em segundos) e a distância euclidiana entre os pontos de partida e chegada. O tempo (em segundos) que demorou para o ônibus chegar na parada é o atributo requerido na previsão.

A Figura 5 apresenta uma amostragem da coleta desses atributos no percurso UFC/IFCE - Praça José de Barros. Nos dados coletados, obtivemos a latitude e longitude do ônibus no momento da coleta, a hora que a localização foi capturada e a velocidade do veículo em Km/h. Observa-se que os dados são gerados, aproximadamente, a cada

10 segundos, o que proporciona uma quantidade de dados satisfatório para a aplicação. Essas informações foram utilizadas no algoritmo de predição.

1	Latitude	Longitude	Hora	Velocidade (Km/h)
2	-4.974.530	-39.034.340	00:40:08	56.06
3	-4.973.767	-39.032.822	00:40:20	57.95
4	-4.973.132	-39.031.536	00:40:30	52.74
5	-4.972.474	-39.030.174	00:40:45	37.21
6	-4.972.107	-39.029.263	00:40:54	46.58
7	-4.971.717	-39.028.191	00:41:03	49.86
8	-4.971.455	-39.027.451	00:41:12	20.72
9	-4.971.333	-39.027.145	00:41:21	16.83
10	-4.971.112	-39.026.531	00:41:30	37.21
11	-4.970.862	-39.025.795	00:41:39	23.39
12	-4.970.808	-39.025.658	00:41:48	0.65
13	-4.970.782	-39.025.635	00:41:57	1.93
14	-4.970.839	-39.025.639	00:42:07	0.07
15	-4.970.850	-39.025.639	00:42:16	1.7
16	-4.970.845	-39.025.642	00:42:25	0.37

Figura 5. Amostragem de coleta no percurso UFC/IFCE - Praça José de Barros

5.3. Resultados e Análises

Nas Figuras 6 e 7 pode-se visualizar, respectivamente, os dados da rota Praça José de Barros - UFC/IFCE e UFC/IFCE - Praça José de Barros. Percebe-se que não foi possível obter uma total cobertura da rota. Isso é ocasionado por dois fatores. O primeiro é o fraco sinal telefônico ao afastar-se do centro da cidade, pois as duas instituições encontram-se distantes do centro da cidade. O segundo motivo é que foi utilizada uma antena para GSM/GPRS com baixo alcance. Apesar dessa limitação inicial do projeto, isso não interferiu no estudo do problema e da viabilidade da solução.



Figura 6. Coleta no sentido Praça (15/03/2019) - UFC/IFCE. Fonte: flopp.net

O módulo GPS utilizado se mostrou bastante preciso, de forma que nenhum ponto foi gerado fora da rota. Este fato contribuiu bastante na coleta e, consequentemente, na predição. Foram realizados diversos testes de predição, com diversos algoritmos e testando empiricamente parâmetros para os mesmos. Também foram testadas



Figura 7. Coleta no sentido UFC/IFCE (15/03/2019) - Praça. Fonte: flopp.net

diversas transformações dos dados, dentre elas a normalização e a padronização. Essas transformações citadas não apresentaram melhores resultados em nossos testes.

Após diversos testes de predição, o gráfico da Figura 8 apresenta uma comparação entre os algoritmos testados. Os testes realizados consistiram em dado o mesmo conjunto de dados de treino para todos os algoritmos, geramos um modelo de aprendizado para cada um dos algoritmos. Após isso, inserimos um mesmo conjunto de dados de teste para todos os algoritmos afim de obter os dados preditos. Por fim, avaliamos os resultados preditos pelos algoritmos utilizando as métricas MAE (*Mean Absolute Error*) e RMSE (*Root Mean Square Error*). A Figura 8 apresenta um resumo dos resultados, onde é exibido os valores das métricas que cada algoritmo obteve nos testes.

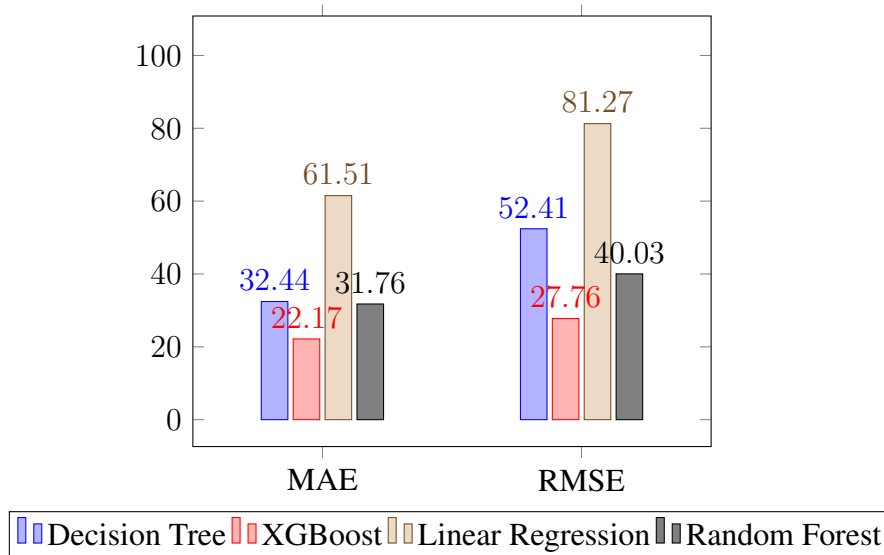


Figura 8. Comparação entre os algoritmos testados utilizando as métricas MAE (*Mean Absolute Error*) e RMSE (*Root Mean Square Error*)

O MAE e o RMSE medem o erro que aconteceu nas predições, dessa forma, o melhor algoritmo em um determinado teste será o que apresentar os menores valores de MAE e RMSE. Pode-se observar diante do gráfico que o algoritmo *XGBoost* foi o que apresentou melhores resultados comparado com os demais algoritmos. Observa-se, tomando

como base a métrica MAE, que em média o *XGBoost* tem erro de aproximadamente 22 segundos, o que é considerado uma precisão muito boa para a aplicação proposta. Pode-se notar que o algoritmo *Linear Regression* foi o que apresentou os piores resultados. Acredita-se que isso se deu pelo fato do tempo não se comportar como uma função linear devido as diversas paradas e o tráfego da via influenciar no tempo.

Os algoritmos *Decision Tree* e o *Random Forest* apesar de apresentar bons resultados não serão utilizados, pois em nenhuma das duas métricas esses algoritmos foram melhor que o *XGBoost*. Diante dos resultados apresentados, foi criado o modelo final utilizando o algoritmo *XGBoost*. Nessa etapa, todos os dados capturados serão utilizados como dados de treino para o algoritmo.

Por fim, foram capturadas informações sobre as *feature importances*, relacionando a importância de cada atributo para a construção do modelo de aprendizado. A *feature importances* apresenta um percentual de importância de um atributo no modelo gerado, ou seja, pode-se observar qual atributo foi mais útil para fazer uma predição mais precisa. Dessa forma, quanto maior o valor para um determinado atributo maior a sua importância para o modelo. A Tabela 1 apresenta os atributos dos dados finais, onde “Dist_euclidiana” é a distância euclidiana entre o ponto de partida e o ponto de chegada, “Tempo” representa o horário que o dado foi capturado e “Velocidade” é a velocidade que o ônibus estava desenvolvendo naquele determinado ponto.

Dist_euclidiana	Tempo	Velocidade
0.28264	0.51920	0.19815

Tabela 1. Importância de cada atributo para o modelo de aprendizado.

Nota-se na Tabela 1 que o tempo é o atributo mais relevante, pois é o atributo que apresenta o maior valor de *feature importances*. Acredita-se que isto está ligado ao fato do tráfego na rota depender diretamente do horário. A velocidade do veículo captada pelo GPS é uma velocidade instantânea. Dessa forma, esse valor pode não ter muita relação com tempo de viagem, já que o veículo pode variar bastante a sua velocidade no intervalo de 10 segundos (que é o tempo para os dados serem capturados). Por esse motivo, acredita-se que a velocidade foi considerada o atributo menos relevante para o modelo de aprendizado final.

6. Conclusão e Trabalhos Futuros

Este trabalho apresentou uma solução baseada em um sistema embarcado e comunicação remota para auxiliar usuários do transporte público ou fretado, especificamente para o transporte de alunos de campi universitários. Adicionalmente uma estratégia para a previsão do tempo de chegada dos ônibus nas paradas foi desenvolvida, baseada em aprendizado de máquina. De maneira geral, a solução atendeu ao objetivo proposto.

Durante o desenvolvimento do trabalho foi analisada a viabilidade do sistema proposto, tanto em custo como em operabilidade. Os testes evidenciaram que em média o algoritmo consegue prever com um erro de apenas 22 segundos, valor baixo comparado ao custo de tempo total de uma viagem (em torno de 30 a 40 minutos).

Como trabalhos futuros pretende-se: implementar o sistema embarcado em um circuito impresso, finalizar o desenvolvimento da aplicação do usuário, adquirir uma

antena com maior alcance e uma placa solar para alimentar o sistema de uma forma ecológica. A partir dessa infraestrutura, pretende-se utilizar as informações sobre o transporte dos universitários em uma aplicação para o usuário, já prevista na arquitetura, concentradas em um *feed* de notícia.

Em relação ao experimento em si, também pretende-se realizar outros trajetos e coletar dados em ambientes totalmente em cidades, com volumes de tráfego e horários variados. Por fim, o desenvolvimento de uma aplicação *mobile* possibilitará ao usuário final obter a previsão de chegada na parada e a localização atual dos ônibus.

O modelo de aprendizado de máquina pode ser aprimorado obtendo outros atributos, tais como, a distância real e o tempo estimado sem paradas que pode ser fornecido pela API (*Application Programming Interface*) do *Google Maps*. A obtenção de mais dados também pode ser encarado como um fator de melhoramento do modelo.

Outra mudança que podemos propor em nosso modelo de aprendizado é a de divisão da rota em diversos trechos de mesmo tamanho e tentar prever o tempo de chegada a partir desses pequenos trechos. Essa abordagem pode ser custosa e não temos garantias de melhorias, mas é apresentada em alguns trabalhos da literatura.

Agradecimentos

Este trabalho foi parcialmente suportado com recursos do projeto **CrOSSiNg: Avaliação da Qualidade de Nuvens Computacionais Apoiadas por Redes Definidas por Software e Virtualização de Funções de Rede**, Edital Universal MCTI/CNPq 01/2016 (processo 422342/2016-5). Agradecemos também ao Programa de Educação Tutorial (PET - Tecnologia da Informação - UFC - Campus Quixadá) pelo apoio no trabalho.

Referências

- Castro, J., Araujo, I., Anjos, E., and Matos, F. (2017). A survey on bus monitoring systems. In *International Conference on Computational Science and Its Applications*, pages 220–231. Springer.
- Cruz, P., Couto, R. S., and Costa, L. H. M. K. (2018). Análise da cobertura espacial de uma rede de sensores baseada em Ônibus urbanos. In *XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC2018)*.
- de M. Neves, A. R., Sarmanho, K. U., Jr., F. C. N., and Meiguins, B. S. (2017). Iniciativa smart campus: um estudo de caso em progresso na universidade federal do pará. In *Workshop de Computação Urbana (CoUrb2017)*.
- Feng, X., Zhang, J., Chen, J., Wang, G., Zhang, L., and Li, R. (2018). Design of intelligent bus positioning based on internet of things for smart campus. *IEEE Access*, 6:60005–60015.
- Martins, N. A. (2005). Sistemas microcontrolados. *Uma abordagem com o Microcontrolador PIC 16F84*. Editora Novatec Ltda, 1ª edição.
- Miranda, W. M., de Mendonça, R. T. R., da Silva, A. A., de Lima Curvello, A. M., de Souza, F. L. d. S., and da Silva, H. J. (2017). Busme: Automatic bus localization system and route registration. *Procedia Computer Science*, 109:1098–1103.

- Monard, M. C. and Baranauskas, J. A. (2003). Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):1.
- Neves, D., Dias, F. C. A., and Cordeiro, D. (2018). Uso de aprendizado supervisionado para análise de confiabilidade de dados de crowdsourcing sobre posicionamento de ônibus. In *Workshop Brasileiro de Cidades Inteligentes (WBCI2018) - XXXVIII Congresso da Sociedade Brasileira de Computação (CSBC 2018)*.
- Rodrigues, D., Severino, J., Costa, F. T., Nakamura, L. H., and Meneguette, R. I. (2018). Uma nova infraestrutura para captação e comunicação dos sensores embarcados no veículo. In *II Workshop de Computação Urbana (CoUrb2018)*.
- Saad, S. A., Ishak, M. H. I., Fauzi, M. H. M., Baharudin, M. A., Idris, N. H., et al. (2018). Real-time on-campus public transportation monitoring system. In *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, pages 215–220. IEEE.
- Sungur, C., Babaoglu, I., and Sungur, A. (2015). Smart bus station-passenger information system. In *2015 2nd International Conference on Information Science and Control Engineering*, pages 921–925. IEEE.
- Vianna, A., Cruz, M., Barbosa, L., and Gama, K. (2018). Análise do impacto de chuvas na velocidade média do transporte público coletivo de Ônibus em Recife. In *Workshop Brasileiro de Cidades Inteligentes (WBCI2018) - XXXVIII Congresso da Sociedade Brasileira de Computação (CSBC 2018)*.
- Wen, M., Rosa, T. O., Souza, M. C., Aleixo, R. P., Silva, C. A., Sá, L. S., Santana, E. F. Z., and Kon, F. (2018). Criação de modelo para simulação de movimentação de Ônibus a partir de dados reais. In *Workshop Brasileiro de Cidades Inteligentes (WBCI2018) - XXXVIII Congresso da Sociedade Brasileira de Computação (CSBC 2018)*.
- Yamaguchi, T., As, M., and Mine, T. (2018). Prediction of bus delay over intervals on various kinds of routes using bus probe data. In *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, pages 97–106. IEEE.