

Construindo mobilidade urbana utilizando tecnologias presentes nos usuários do transporte público através do algoritmo J48

Lucas Alves São Mateus

*Aluno do Curso Ciência da Computação – Universidade Federal do Pampa (UNIPAMPA) Alegrete – RS
– Brasil*

DOI: 10.47573/aya.5379.2.94.20

RESUMO

Mobilidade deveria ser um direito universal. Porém, os conceitos de cidade e mobilidade não estão completamente em harmonia. No Brasil, boa parte da vida dos trabalhadores é desperdiçada no trânsito caótico das grandes cidades. O transporte público está longe de ser eficiente e o poder público está aquém desse desafio. Diante disso, vários trabalhos propõem soluções tecnológicas em busca da mobilidade urbana com foco na previsão do tempo de viagem em tempo real. Nos últimos anos, soluções eficientes surgiram (p.ex. Moovit) utilizando redes móveis e smartphones através de vários algoritmos que possibilitam a previsão com acurácia: Rede neural artificial, filtro de Kalman e regressão linear. Porém, ainda assim, cidades brasileiras estão longe de prover informações simples, como os horários previstos de uma rota. Com o objetivo de atender o usuário de transporte público, propõe analisar a eficiência do J48 para identificação do tipo de transporte, usando somente acelerômetro, sendo utilizado pelo usuário, evitando uso do GPS para evitar invasão da privacidade do usuário.

Palavras-chave: J48. transporte público. algoritmos. mobilidade.

INTRODUÇÃO

Cidade e mobilidade são indissociáveis. Porém, no Brasil, a política do transporte público é insuficiente e insustentável. Além disso, nos últimos anos, o transporte individual teve um aumento significativo - o exemplo maior aconteceu na cidade do Rio de Janeiro cuja frota, em apenas 12 anos, cresceu 73%. Esse fenômeno provocou uma mobilização urbana generalizada (PAULA e BARTELT, 2016).

Um dado que ilustra essa imobilização, é o fato do aumento do tempo médio de deslocamento em dez regiões do Brasil (de 38,1 para 43,3 minutos) nos últimos dez anos. Apesar do cenário negativo, os autores Paula e Bartelt (2016) afirmam que os usuários de transporte individual usariam os meios de locomoção públicos, se os mesmos, fossem de qualidade.

Diante dos problemas levantados, este trabalho irá contribuir no seguinte sentido: demonstrar a eficiência do J48 na classificação das atividades (correr, andar de ônibus e etc) dos usuários. Para alcançar esse objetivo, foi aproveitado tecnologias presentes no dia a dia dos passageiros, como smartphones e redes móveis. Portanto, sem a necessidade de intervenção física (como ocorreu em Porto Alegre-RS (LADEIRA *et al.* 2011). Demonstrando que é possível, alcançar um direito humano: Mobilidade Urbana.

Os celulares dos usuários são múltiplas fontes de dados que podem ser transformados em informações cruciais para atender com eficiência o transporte coletivo - intensidade do tráfego, velocidade média, previsão da hora de chegada e outras informações que podem ser fornecidas sem a criação de uma nova estrutura, somente usando as redes móveis.

Ao observar as tecnologias presentes e com baixo custo de intervenção se faz necessário aprofundar em temas, que podem tornar a vida de milhares de usuários do transporte público mais dignas.

OBJETIVO

O objetivo deste trabalho é analisar a eficiência do algoritmo J48, um algoritmo capaz de realizar a classificação da atividade do usuário sem utilizar dados de GPS. Dado o objetivo geral, será necessário alcançar os seguintes objetivos específicos:

- Identificar as características básicas que compõem o transporte público, com objetivo de verificar as condições que podem atrair o usuário para o transporte público;
- Selecionar, a partir de uma revisão, projetos com o mesmo objetivo;
- Selecionar o algoritmo J48 e adicioná-lo a uma aplicação no Android Studio;
- Realizar a análise dos resultados de classificação a fim de verificar se o desempenho corresponde à necessidade do projeto;

METODOLOGIA

Foram utilizadas fontes científicas disponíveis via web, para realizar a busca por artigos de periódicos publicados nos últimos anos, sendo aplicadas as palavras-chave previamente definidas. Na tabela 1 são mostradas as palavras em português com suas traduções na língua inglesa.

Foram utilizadas as seguintes bases de pesquisa:

- IEEE Explore <<http://ieeexplore.ieee.org>>;
- SCOPUS <<https://www.scopus.com/home.uri>>;
- Google Scholar <<https://scholar.google.com.br/>>

Tabela 1- Palavras-chave utilizadas na string de busca

Frases chaves	Sinônimo em Inglês
predição de posição veicular	vehicle position prediction
algoritmos predição de posição	position prediction algorithms
j48 predição de posição	j48 position prediction

A seguir os Critérios de Inclusão:

1. Critério 1: devem ser trabalhos publicados e disponíveis integralmente nas bases de dados definidas no trabalho;
2. Critério 2: Somente artigos em inglês;
3. Critério 3: os trabalhos devem tratar sobre o predição de posição de veículos

A seguir os Critérios de Exclusão:

1. Critério 1: serão desconsiderados trabalhos que não estejam disponíveis integralmente nas bases de dados utilizados;

As buscas com base nas palavras chaves foram executadas nos bancos de dados de-

finidos. Os trabalhos selecionados, foram lidos os resumos e uma pré-avaliação com base nos critérios escolhidos. Os textos escolhidos foram avaliados através da leitura integral sendo considerados válidos ou não para o objetivo deste trabalho.

Para obter as informações foram preenchidos para cada texto escolhido, um "formulário de extração de dados", com objetivo de sistematizar o grau de importância da contribuição para este trabalho. Além dos dados básicos (dados bibliográficos, data de publicação, abstract, contexto, objetivo, metodologia e resultados).

RESUMOS DOS TRABALHOS

Foram dois campos de pesquisas, o primeiro para compreender o estado atual do transporte público brasileiro e seu potencial. O segundo, visando compreender as técnicas utilizadas para oferecer serviços cruciais para os usuários do transporte público.

Transporte público

As cidades brasileiras estão distantes do conceito de cidade inteligente. O trabalho Paula e Bartelt (2016) comenta que o recurso destinado ao sistema de transporte inteligente tem como foco o controle dos cidadãos, isso fica evidente no direcionamento dos recursos na compra de câmaras de HD, que são usadas nos grandes centros de controle.

Contudo, esse investimento não é direcionado de forma positiva para o usuário, principalmente na questão de tempo no transporte público, visto que o deslocamento corresponde a mais de 10% do seu dia (PAULA e BARTELT, 2016).

Diante disso, o autor Singh *et al.* (2017) comenta que uma das características que o usuário mais preza é a previsibilidade do transporte público. Além desta, o conforto foi outro aspecto levantado. Para que ocorra a migração do transporte individual para o coletivo é necessário que esse item (conforto) seja adotado.

Os direitos humanos estão em confronto direto com o transporte público brasileiro, visto que não há qualidade no transporte público, violência, sem previsibilidade. Ao ponto que a empresa Moovit, uma empresa israelense de mobilidade, afirmou que inúmeros trabalhadores da cidade de Porto Alegre passam 3.5 horas por dia nos deslocamentos da sua casa até o trabalho.

Pesquisas sobre a predição de posição veicular

O trabalho Singh *et al.* (2017) defende que o desempenho dos algoritmos de predição estão relacionadas às características de trânsito, como por exemplo: se os motoristas são disciplinados, semáforos, interseções e outros. Essa afirmação é comprovada através de comparações entre os algoritmos de regressão, FK e Rede Neural Artificial em dois cenários: Países desenvolvidos e em desenvolvimento.

Primeiro, essas soluções já são uma realidade com a aplicação de alguns algoritmos com desempenho satisfatório. Já no segundo, devido à heterogeneidade do trânsito, a precisão dos algoritmos falha no objetivo de oferecer serviços de predição de posição aos usuários. De acordo com os resultados obtidos por Singh *et al.* (2017), é constatado que os algoritmos têm

sua eficiência alterada de acordo com o trânsito da região.

Na construção da solução deste trabalho, o artigo Gurmu (2014) contribuiu através das comparações entre a Média Histórica (MH), Filtragem de Kalman (FK) e Rede Neural Artificial (RNA) usando o modelo MAPE (Mean Absolute Percentage Error) para definir qual teria a maior eficiência.

De acordo com o Gurmu (2014) vantagem do modelo MH é que o tempo de computação é relativamente pequeno, todavia, o desempenho não é satisfatório. Já o FK possui a capacidade de filtrar ruídos, o que é fundamental para prever a precisão dos veículos. Já o RNA, além de filtrar ruídos, é capaz de lidar com dados complexos e relacionar variáveis dependentes com independentes. Apesar da eficiência do RNA, o autor fez comentários mencionando que as condições do tráfego influenciam nos resultados. Além disso, observou a necessidade de um banco de dados suficiente para aplicação do RNA.

Os resultados obtidos pelo MAPE para verificar o desempenho foram divididos em três tipos de trajeto: curto, médio e longo. O algoritmo que apresentou o melhor desempenho foi RNA. No quesito curto, os três algoritmos apresentaram praticamente os mesmo resultados, contudo, em alguns momentos MH apresentou melhores resultados que o RNA. I

O autor Sun *et al.* (2007) propôs uma solução mais vantajosa que MH, através das relações das variáveis independentes, como o tempo médio das paradas dos ônibus e a posição dentro do segmento. Além dessa contribuição, utilizou uma máquina de estados finitos para evitar o problema backward data. O desempenho do algoritmo proposto por Sun *et al.* (2007) teve um resultado com erro menor de 5%.

Todos os modelos de previsão vistos anteriormente utilizam dados de GPS, porém o trabalho de Zhou *et al.* (2012) propôs uma alternativa visando proteger a privacidade do usuário e evitar custo de energia envolvendo as operações com GPS.

O autor Zhou *et al.* (2012) propôs utilizar as torres de celulares e os áudios coletados dos smartphones e dados dos sensores para obterem posição geográfica do veículo e a previsão do tempo da viagem.

O primeiro passo da proposta Zhou *et al.* (2012), foi identificar quando o usuário estará dentro do veículo. A solução foi usar o som emitido pelo aparelho mostrado na figura 1 quando utilizado pelos usuários. O resultado foi uma precisão de 95%, considerando a captação do som a cerca de três metros. Além dessa identificação, o autor mostrou ainda que é possível identificar quais dos passageiros estão compartilhando o mesmo ônibus, através dos intervalos dos áudios emitidos pelo aparelho da figura 1.

Contudo, o aparelho da figura 1 é utilizado para outros serviços, como trens, provocando falsos positivos para identificação do transporte. Para resolver esse impasse, o autor Zhou *et al.* (2012) usou o acelerômetro do smartphone para distinguir esses tipos de transporte. O comportamento da aceleração do ônibus possui características que possibilitam sua identificação. A precisão do resultado foi de 90%.

Figura 1 - Cartão de barramento IC



Aparelho que emite som

Figura 2 - Trajeto do Ponto A ao B



Segmentos dos ônibus através das torres de celular

O autor Zhou *et al.* (2012) demonstra que é possível identificar o segmento pelas torres do celular. A figura 2 demonstra o deslocamento do ponto A até o ponto B. O autor Zhou *et al.* (2012), usou a seguinte estratégia: um banco de dados com três torres de maior intensidade para cada segmento. Ao comparar o banco de dados com informações das torres conectadas pelo smartphone (7,8,4,5) da figura 2 seria possível determinar qual o segmento atual do ônibus, com base no histórico das torres conectadas. Porém, essa simples consulta do banco de dados, para cenários complexos, mostrou que essa abordagem não foi eficiente.

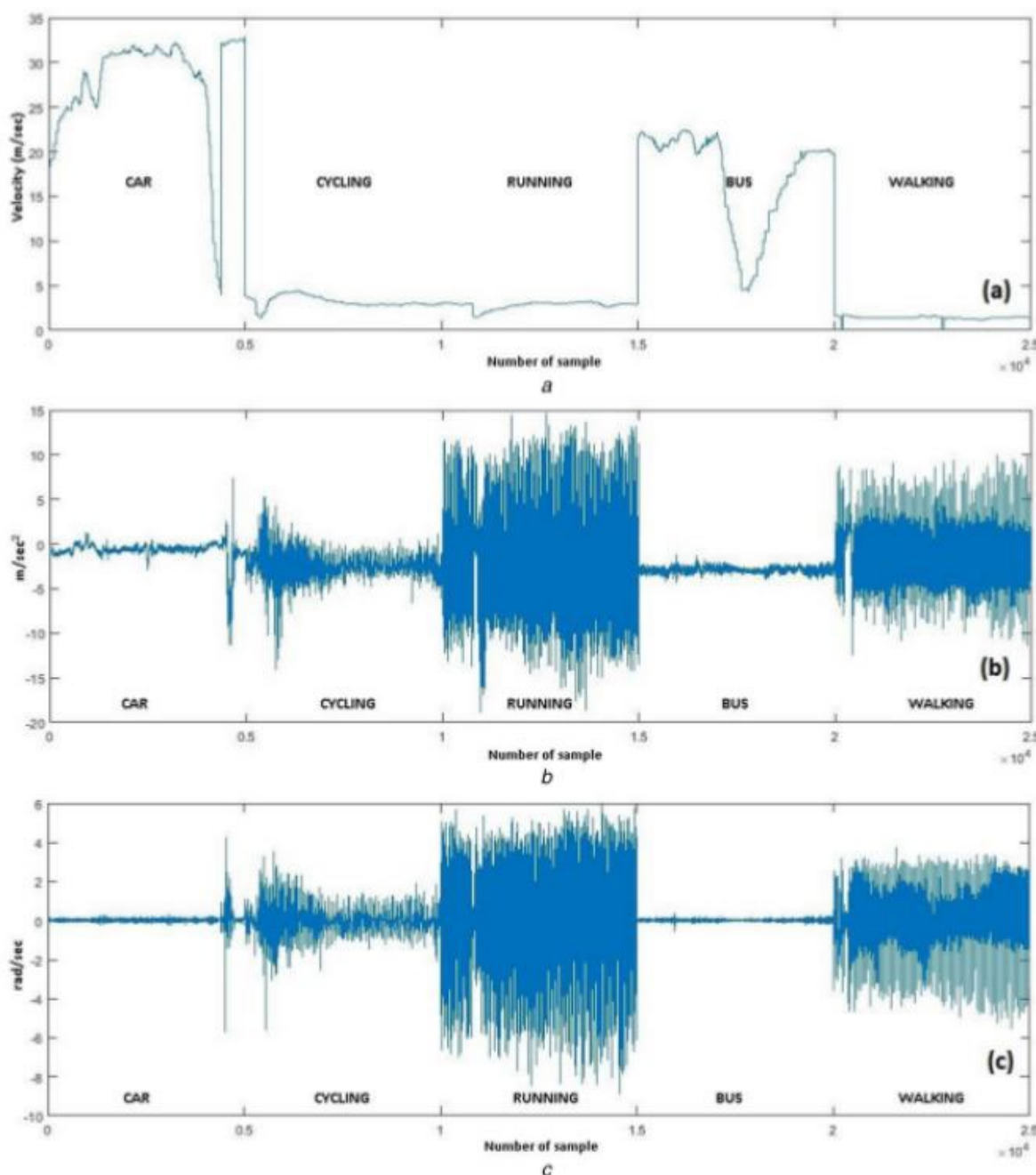
Os resultados obtidos pelo trabalho Zhou *et al.* (2012) sobre detecção do tipo de transporte poderia ter precisão maior, caso tivesse usado a combinação de sensores presente em um smartphone. O trabalho Balli e Sagbas (2017) aplicou essa possibilidade e obteve resultados expostos a seguir.

Os autores se concentraram na identificação de cinco tipos de locomoção: caminhar, correr, andar de bicicleta, carro e ônibus. Para isso, utilizou dados de acelerômetro, giroscópio e sensores de GPS coletados através de um aplicativo em celular. A figura 3 demonstra que cada tipo de locomoção tem suas características em relação aos tipos de dados coletados. Em relação

ao GPS, cada tipo de transporte têm velocidades médias diferentes, porém, em condições de tráfego intenso podem apresentar as mesmas velocidades.

O acelerômetro e o giroscópio conseguem distinguir facilmente transportes não motorizados. Em relação ao carro e ônibus, embora se vejam semelhanças entre eles, as velocidades angulares são diferentes.

Figura 3 - Amostra de dados do sensor



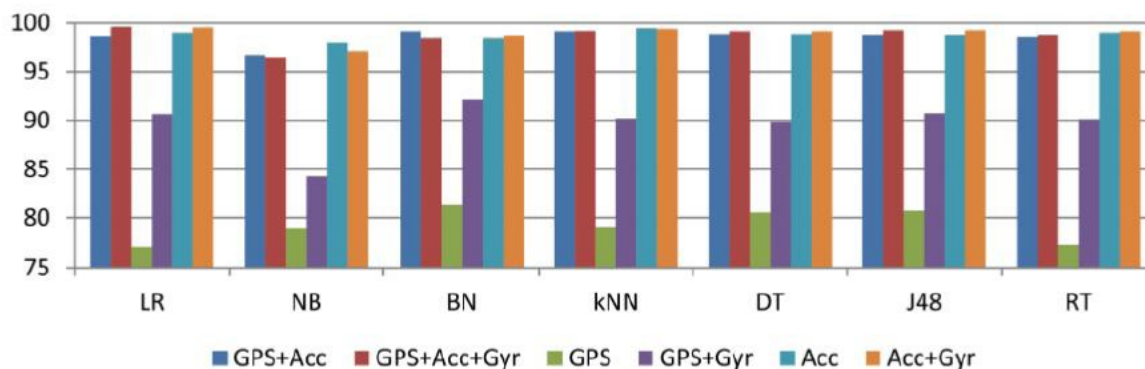
(a) Velocidades - (b) Acelerômetro(eixo-X) -(c) Giroscópio(eixo-Y)

Ballı e Sagbas (2017)

Com os dados coletados, o autor aplicou alguns métodos supervisionados de aprendizado de máquina: Bayes Ingênuo(NB), rede Bayesiana(BN), K- nearest neighbour(KNN), regressão logística(LR), J48, tabela de decisão(DT) e árvore aleatória(RT).

Os resultados do trabalho Ballı e Sagbas (2017) na figura 4 mostram que a combinação dos três sensores e o método de regressão logística alcançaram uma precisão de 99,6%. Contudo, a precisão utilizando somente o acelerômetro e o giroscópio chegou a 99,5%, o que é um resultado considerável, devido ao consumo de energia das operações do GPS e questões sobre privacidade do usuário.

Figura 4 - Taxas de precisão de combinações dos sensores



Ballı and Sagbas (2017) Acc (acelerômetro) e Gyr(giroscópio)

PROPOSTA DO TRABALHO

Identificar a precisão do algoritmo J48, utilizando o para a determinação da atividade que está sendo executada (correr, andar, ônibus e etc...). Além disso, identificar a relação entre a quantidade de informações fornecidas e a sua precisão. Para esse fim, cenários de testes serão criados. Vale ressaltar, que para predição da posição do veículo ocorra é crucial que a precisão de identificação seja extremamente eficiente.

Para alcançar a proposta, é utilizado um conjunto de dados produzidos pela Universidade Sussex Huawei Locomotion. Os dados coletados através dos sensores do celular tendem a representar o comportamento do indivíduo ao realizar uma atividade. O conjunto de dados disponibilizados contém 3 usuários. O total de dados computados para análise foram 27,4 GB. Os dados utilizados têm um diferencial que é a coleta realizada por quatro smartphones colocados em locais diferentes no corpo do usuário, ou seja, possibilitando uma maior aproximação com a realidade dos usuários.

Para atingir a proposta, foram realizados 45 testes. Dentro de cada proporção dos dados utilizados, foram utilizadas as seguintes médias: 500, 400, 300, 200 e 100. Cada média foi testada com 6 tipos de locomoção.

As seis classes são constituídas pelas atividades para ficar parado, andando, correndo, bicicleta, carro e ônibus.

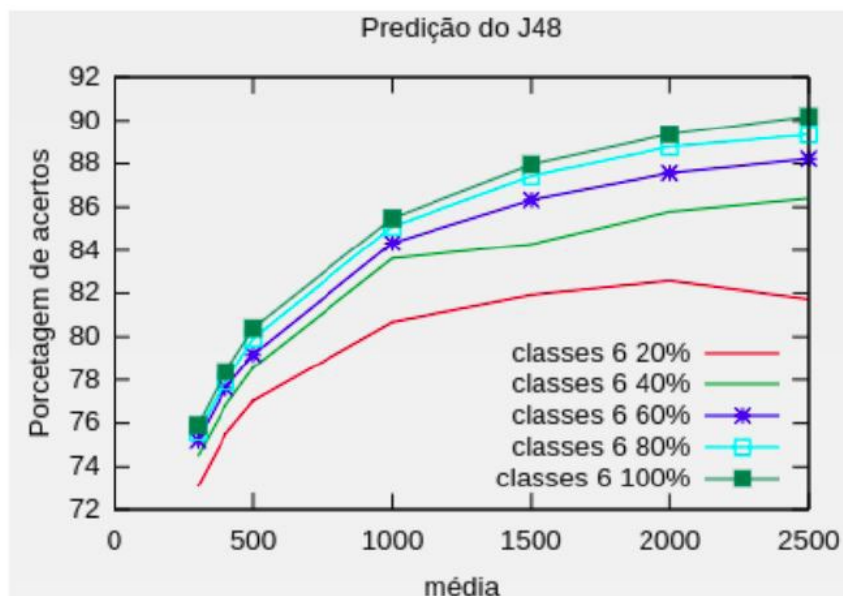
Para atingir o objetivo deste trabalho foi usado a ferramenta Waikato Environment for Knowledge Analysis - WEKA de classificação através de do algoritmo J48. Não será analisado o funcionamento do J48 neste artigo, somente suas capacidades de determinar qual atividade do usuário.

RESULTADOS

Neste capítulo serão apresentados os resultados obtidos com os 45 testes realizados para predição da atividade utilizando o J48. Foram avaliados os impactos do número de elementos utilizados na média, inclusive no momento diminuição da taxa de predição, taxa de falsos positivos (FPR), juntamente com a matriz de confusão.

Cada atividade possui um comportamento, que é expresso na variável escolhida - acelerômetro. A figura 5 demonstra que o comportamento da curva é semelhante a uma função $F(x) = \log(x)$. Outro dado é que quando há um aumento do número de elementos na média, há uma maior porcentagem de acertos do J48. Observe na figura 5, que com amostragem dos dados de 60% (linha roxa) a correta classificação das instâncias aumenta com o número de elementos utilizados na média - (nº1000 - 73.37%), (nº2000 - 76.27%) até um ponto crítico.

Figura 5 - Figura 1



Pode-se analisar o ponto crítico ao observar a figura 8, que compara o ganho de acertos em relação à média anterior. A taxa de acertos está crescendo a cada aumento do número de elementos usados na média (nº 500 para no 1000 elementos aumentou 3,64% a taxa de acertos). Para exemplificar esses números, a precisão na figura 7 com 500 elementos é de 77,04%, já com 1000 elementos é de 80,68%, ou seja, a diferença é de 3,64%. Esses dados são importantes para encontrar o ponto crítico, que nesse ocorreu com 2500 elementos, onde a taxa de acertos decaiu para -0,86%. Vale ressaltar que esse comportamento está relacionado ao conjunto de dados fornecidos ao J48. Além disso, todos os testes tendem a ir ao encontro deste ponto, como fica claro na figura 6, com aumento no eixo "x" há uma diminuição da taxa de ganho em relação ao número de elementos na média anterior.

No trabalho Ballı e Sagbas (2017) não analisou duas diferenças que impactam no resultado da acurácia deste trabalho. O primeiro fator, as 4 posições do celular. Não há referências no trabalho, Ballı e Sagbas (2017), sobre esse fato. A segunda, é a atividade parada, ou seja, dentro de um prédio, casa e etc. O J48 encontrou dificuldades com essa atividade, ao ponto que apresentou o maior taxa de falso positivo (FPR), como consta na tabela 2. Ou seja, na maioria das classes a atividade "parado" impactou no resultado.

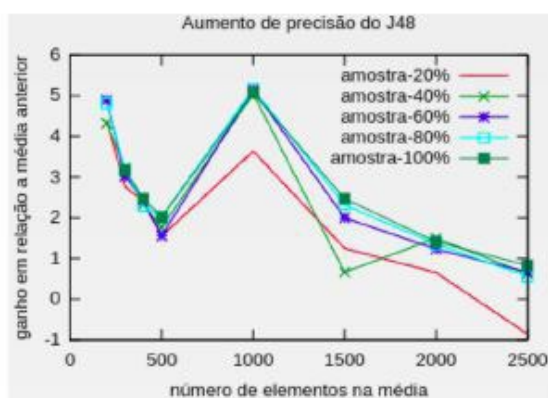


Figura 6 - Figura 1



Figura 7 - Figura 2

Figura 8. Fonte: Elaborado pelo autor.

Classes	Parado	Caminhando	Correndo	Bicicleta	Carro	Ônibus	FPR
Parado	5238	74	0	171	190	186	0,034
Caminhando	66	5453	2	123	39	25	0,015
Correndo	0	2	580	0	0	0	0,000
Bicicleta	195	136	0	3687	91	54	0,023
Carro	149	31	0	95	4001	185	0,028
Ônibus	195	27	0	59	213	2275	0,022

Tabela 2

Essa atividade ao observar a tabela 2 na coluna Falsos positivos(FPR) apresentou um resultado acima da média(soma de todas FPR dividido por 5) das outras atividades. Com esse dado, pode se afirmar que a mesma provocou o maior impacto na taxa de acurácia do algoritmo.

Sobre a tabela 2, a matriz de confusão foi construída com 100% dos dados disponíveis. A escolha de demonstrar somente essa matriz de confusão foi devido, que a mesma apresentou os melhores resultados. Em relação ao número de elementos utilizados na média foi de 2500. Observe que os maiores números estão na diagonal, o que mostra a precisão do algoritmo. Nas atividades, caminhando e correndo, foram as que representaram as maiores precisões com 5453 e 580 acertos, respectivamente. Esses números representam uma acurácia de 95,3% e 99,7% - tabela 3. Essas duas atividades apresentaram os FPR resultados significativos, 0,015 e zero(0.0).

No que diz respeito `a atividade "correndo", apesar dos limites dos dados, o J48 determinou com uma acurácia significativa. Ao observar a figura 3 fica notório a diferença dos dados do acelerômetro para outras atividades, exceto a parada. Por isso, o J48 apresentou, para essa atividade, o seu melhor resultado. Apesar da limitação dos dados não foi um fator limitador.

Sobre as outras quatro atividades, todas apresentaram, em todos os cenários, resultados menores de 90%, como demonstrado na tabela 3. Contudo, todas as atividades aumentaram sua precisão, quando o número de elementos para cálculo da média sofreu alteração. Ao ponto da atividade ônibus iniciar com 66% e terminar com 83%.

Classes	*Nº elementos para o cálculo da média				
	500	1000	1500	2000	2500
Parado	80,30%	84,60%	87,10%	88,10%	89,60%
Caminhando	90,70%	93,40%	94,50%	94,80%	95,30%
Correndo	99,60%	99,70%	99,70%	99,50%	99,70%
Bicicleta	77,90%	84,10%	86,80%	88,10%	89,20%
Carro	75,50%	82,80%	85,60%	97,80%	88,20%
Ônibus	66,80%	74,50%	79,40%	82,90%	83,50%
Precisão Geral	80,40%	85,50%	87,90%	89,30%	90,20%

Tabela 3 - Acurácia por Classe

A atividade de ônibus, foco deste trabalho, apresentou resultados relevantes. O primeiro, é que sua precisão alcançou 83%, sendo possível identificar uma relação de ganho entre precisão e número de elementos usados na média. O segundo, é em relação FPR, pois apresentou resultado inferior - 0,022 - a média dessa variável que é 0,024. Os resultados usando o J48 sofreram impactos devido `as altas taxas de falsos positivos das classes - bicicleta(54), carro(185) e "parado"(186).

Sobre a detecção da atividade carro, demonstrou que é possível alcançar resultados importantes, como a precisão 88,2% na última coluna da tabela 3. Porém, encontrou os mesmos impasses dos falsos positivos nas atividades "parado", ônibus e bicicleta. Todavia, com melhores resultados.

Já atividade bicicleta encontrou os mesmos impasses, nas últimas duas atividades citadas, porém com resultados melhores - 89,2%.

Os dados do sensor acelerômetro mostraram que pode ser usado com precisão considerável para classificar qual é atividade que está sendo executada.

CONSIDERAÇÕES FINAIS

Buscando fomentar novos trabalhos e buscando criar uma relação de conhecimento com mudança de vida população, este trabalho teve como objetivo geral a demonstração da eficiência do J48 para classificação de qual atividade está sendo executada com objetivo de determinar se um indivíduo está usando o transporte público.

O grande achado deste trabalho é o impacto da atividade "parado" nos resultados dos testes. Os valores do FPR são altíssimos. Conclui-se que o J48 encontrou dificuldades para classificar essa atividade. Uma solução para essa situação é encontrar adicionar novos dados, de outros sensores buscando contornar essa situação - como pode ser giroscópio.

Ao analisar os resultados dos testes, foi possível identificar que a uma relação entre a amostragem utilizada na média e a acurácia. Utilizando como referência a tabela 2 conclui que a amostragem da média deve ser dinâmica, pois há um aumento na acurácia quando se modifica a amostragem da média.

Ao adentrar na acurácia do J48 deste trabalho, o J48 apresentou resultados significativos para seis classes - 90.2% de acurácia geral. Para trabalhos futuros, pretende-se utilizar a

validação da proposta com outras variáveis e adicionando outros algoritmos nos testes.

A contribuição deste trabalho está no fato de que é possível utilizar ferramentas já disponíveis dos usuários do transporte para oferecer um serviço de qualidade. O direito da mobilidade nas cidades brasileiras já está nas mãos dos passageiros, seus smartphones.

AGRADECIMENTOS

Gostaria de agradecer ao professor Marcelo Resende Thielo pela oportunidade de trabalhar neste estudo.

REFERÊNCIAS

BALLI, S. E SAGBAS, E. A. (2017). Diagnosis of transportation modes on mobile phone using logistic regression classification. *IET Software*, 12(2):142–151.

Gurmu, Zegeye Kebede e Fan, W. D. (2014). Predicting bus arrival time on the basis of global positioning system data. Technical report.

LADEIRA, M., Michel, F., e Senna, L. (2011). Public transport monitoring and control: The case of porto alegre, brazil. In *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation*, pp. 275–281.

PAULA, M. D. e BARTELT, D. D. (2016). Mobilidade urbana no brasil: desafios e alternativas. Rio de Janeiro: Fundação Heirich Boll.

SINGH, G., BANSAL, D., e SOFAT, S. (2017). Eta htc: Estimating time of arrival under heterogeneous traffic conditions using crowdsensing. In *2017 International Conference on Inventive Computing and Informatics (ICICI)*, pages 175–179. IEEE.

SUN, D., LUO, H., FU, L., LIU, W., LIAO, X., e Zhao, M. (2007). Predicting bus arrival time on the basis of global positioning system data. *Transportation Research Record: Journal of the Transportation Research Board*, (2034):62–72.

ZHOU, P., ZHENG, Y., e LI, M. (2012). How long to wait?: predicting bus arrival time with mobile phone based participatory sensing. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 379–392. ACM.