

Classificação da Qualidade de Vias Urbanas baseado em Sensoriamento Participativo*

Davidson Esteves Nunes¹, Amanda Fagundes¹, Vinícius F. S. Mota^{1,2}

¹ Departamento de Computação e Sistemas
Universidade Federal de Ouro Preto (UFOP)
João Monlevade – MG – Brasil

² Departamento de Informática
Universidade Federal do Espírito Santo (UFES)
Vitória – ES – Brasil

davidson@gmail.com, amanda.fagundes@ymail.com, vinicius.mota@inf.ufes.br

Abstract. *Besides the number of initiatives for real-time traffic monitoring, fewer of them bring asphalt quality information, an important aspect in the route decision process. We propose a framework, called StreetCheck, for participatory sensing and, further, urban roads classification. Streetcheck gathers mobile devices sensors information such as GPS and accelerometer, and also user opinion about the quality of the road. Based on the data gathered, we mapped a set of features as input for supervised learning algorithms. To validate Streetcheck, 20 volunteers drove over 1200 km of urban roads. Streetcheck reached up to 90,64% of hit rate.*

Resumo. *Embora existam diversas iniciativas para informar o tráfego em tempo real, poucas trazem informações sobre a qualidade da via no trajeto, aspecto que pode ser decisivo na tomada de decisão entre rotas. Este trabalho propõe um arcabouço, denominado Streetcheck, para sensoriamento participativo e, posteriormente, classificação de vias urbanas. O Streetcheck coleta informações como GPS, acelerômetro e opinião sobre a via em dispositivos móveis. A partir dos dados coletados, foram mapeados um conjunto de características para servirem como entrada para algoritmos de aprendizado supervisionados. Para validar nossa proposta, 20 usuários percorreram aproximadamente 1200 km. O Streetcheck alcançou taxas de acerto na classificação de vias de até 90,64%.*

1. Introdução

O transporte rodoviário é o mais utilizado no Brasil e possui uma ampla representatividade nos seus custos logísticos. No entanto, um dos grandes problemas enfrentados pelos brasileiros é a estrutura rodoviária [CNT 2016]. Tal situação é apontada pela CNT como um dos fatores do elevado índice de acidentes e incidentes nas ruas e rodovias brasileiras.

Uma maneira dos motoristas tentarem amenizar este problema é obtendo informações sobre situação do trânsito e qualidade da via em tempo real, adequando suas

*Os autores agradecem à UFOP, UFES e à FAPES pelo apoio financeiro parcial a este trabalho.

rotas. Uma forma de obter estas informações é por meio do sensoriamento participativo ou *Crowdsensing*. O sensoriamento participativo é caracterizado como coleta de dados de forma coletiva que permite gerar informação agregada útil a todos os participantes [Burke et al. 2006]. Um dos grandes desafios é incentivar os usuários a colaborarem fornecendo dados, muitas vezes, privados. Porém, o usuário pode levar em consideração se a vantagem da informação agregada é maior que o ônus da colaboração [Silva et al. 2016]. De fato, aplicativos como *Waze*¹ e *MapLink*² fornecem informação sobre tráfego em tempo real baseado em dados de seus usuários.

No entanto, a qualidade da via pode ser decisiva na tomada de decisão de uma rota, uma vez que o motorista poderá optar por aquela que possua uma avaliação superior. Considerando o fato de que tais vias podem conter buracos, lombadas e asfalto irregular, o processo de classificação da qualidade de vias urbanas se mostra desafiador.

Neste contexto, este trabalho propõe um arcabouço, denominado *Streetcheck* para coleta de informações, transmissão e classificação de vias urbanas. O objetivo do *Streetcheck* é caracterizar, utilizando aprendizado supervisionado, trechos completos de vias urbanas. O *Streetcheck* coleta dados dos sensores de geolocalização e acelerômetro de dispositivos móveis. Além disto, permite que os usuários opinem sobre a qualidade de um determinado trecho. As informações coletadas são encaminhadas para serem utilizadas em um sistema de classificação.

O sensor de geolocalização fornece a posição geográfica e velocidade do dispositivo. Já o sensor acelerômetro fornece a aceleração, em três dimensões (x, y, z) , exercida sobre o mesmo. O dispositivo dentro do veículo receberá as vibrações que passam da via para o interior do veículo. No entanto, as leituras de acelerômetro dependem da posição dentro de um veículo, da qualidade da suspensão, entre outras. Para contornar este problema, foram extraídos um conjunto de características que serviram de entrada para algoritmos de aprendizado supervisionado.

Um grupo de vinte voluntários percorreram 1.218,34km durante 3 semanas, opinando (rotulando) sobre a qualidade da via, quando possível. A partir dos dados coletados, foram avaliadas as relevâncias das características e quais algoritmos de aprendizado supervisionado tem melhor desempenho no problema de classificação de vias. As contribuições deste trabalho são sumarizadas a seguir:

- Propõe um arcabouço para sensoriamento participativo e inferência da qualidade das vias;
- Propõe um conjunto de características para mensurar a qualidade de vias;
- Demonstra uma prova de conceito da solução, com um número pequeno de usuários mas com uma expressiva quantidade de dados.

O restante do artigo está organizado como se segue: A Seção 2 apresenta os trabalhos relacionados. A arquitetura do arcabouço, assim como as funções de mapeamento dos sensores em características para algoritmos de aprendizado supervisionados são descritos na Seção 3. A metodologia dos testes e os dados coletados são descritos na Seção 4. Os resultados obtidos são discutidos na Seção 5. Por fim, a Seção 6 conclui e discute os trabalhos futuros.

¹www.waze.com

²transito.maplink.global

2. Trabalhos Relacionados

Um dos primeiros trabalhos para a detecção de irregularidades em vias baseado em acelerômetro e posicionamento por satélite foi proposto em um sistema de [Eriksson et al. 2008] batizado *Pothole Patrol P²* para detectar irregularidades nas vias. Para isto, utilizaram um acelerômetro nos painéis de um conjunto de táxis em Los Angeles e um algoritmo *Naive* para classificar as irregularidades. Uma outra proposta baseada em heurísticas de limites de leitura dos sensores é apresentada em [Mednis et al. 2011]. Os autores conseguem uma taxa de acerto de até 92% na detecção de anomalias.

Em [González et al. 2017], os autores propõem o uso da abordagem *Bag of words* para criar o vetor de características como entrada para algoritmos de aprendizado supervisionado. Embora os algoritmos apresentados tiveram resultados estatisticamente similares, o algoritmo de rede neural apresentou melhor desempenho. As medições foram feitas com o *smartphone* em diversas posições no veículo de forma a verificar a influência deste fator nos dados. Neste projeto também esse fator foi levado em conta, de forma que o usuário pode registrar o posicionamento do dispositivo durante o uso.

Utilizando sensoriamento participativo, uma arquitetura batizada de RoadScan [Lima et al. 2016], visa classificar vias utilizando como métricas o desvio padrão das leituras do acelerômetro e algoritmos de detecção de picos.

Com exceção do RoadScan, os demais trabalhos focam na detecção de anormalidades como buracos, lombadas ou junções de pistas. No entanto, as vias brasileiras apresentam particularidades como as ondulações, que podem passar despercebidas e comprometer o desempenho dessas ferramentas. Neste sentido, este trabalho propõe um arcabouço para classificação de trechos urbanos completos utilizando algoritmos de aprendizado supervisionado. Para isto, foram extraídos um novo conjunto de características na tarefa de classificação das vias, além de permitir que usuários opinem sobre a qualidade das vias.

3. Arquitetura do Streetcheck

3.1. Visão Geral

O sistema para a classificação de vias é dividido em duas partes: (i) **Coleta de dados**, implementado como um aplicativo que pode ser instalado em dispositivos móveis; e (ii) **Sistema de classificação**, implementado em servidor em nuvem, responsável por receber os dados, armazená-los e posteriormente, filtrar, mapear e aplicar os dados nos algoritmos de classificação.

A arquitetura do *Streetcheck* e as etapas de processamento dos dados são apresentadas na Figura 1. O *Streetcheck* considera como fonte de dados dispositivos móveis equipados com sensores de geolocalização, acelerômetro e interface com usuário. Deste modo, o sistema de coleta de dados captura periodicamente as informações desses sensores e também, a opinião do usuário sobre a qualidade da via. Em seguida, um pré-processamento sumariza os dados e remove leituras errôneas.

Após serem transmitidos para um servidor, os dados brutos são filtrados e as características são extraídas. As amostras rotuladas pelos usuários alimentam os algoritmos de aprendizado supervisionados, já as amostras não rotuladas são submetidas ao processo

de predição. Com as vias classificadas, um usuário pode ter uma visão da qualidade das vias que outros usuários já percorreram.

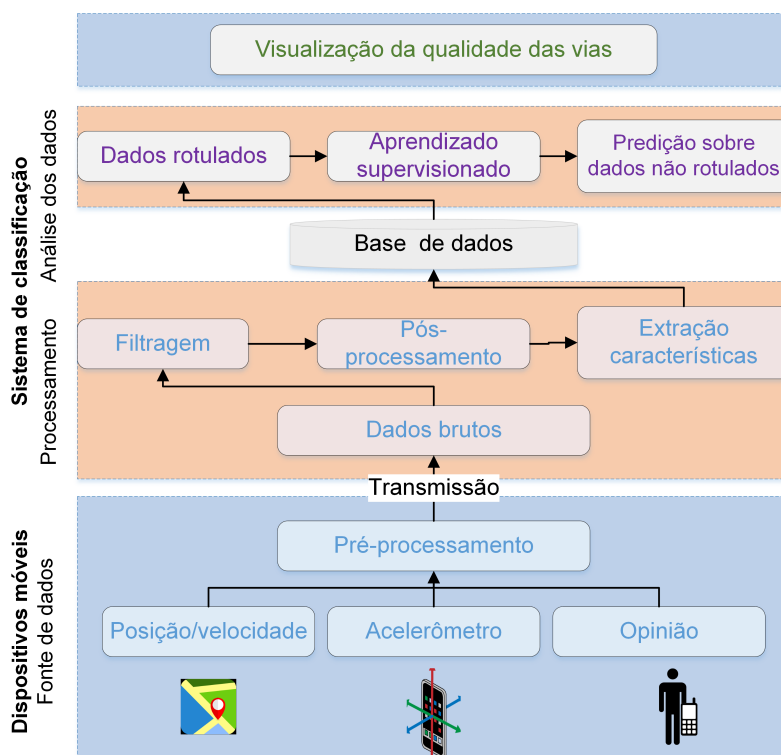


Figura 1. Arquitetura do sistema

3.2. Sistema de coleta de dados

O sistema de coleta de dados deve ser executado nos dispositivos móveis dos usuários. O *Streetcheck* coleta periodicamente leituras realizadas pelo acelerômetro do dispositivo e à localização do usuário, além de permitir que o usuário rotule um trecho da via em que esteja passando.

O sensor acelerômetro fornece a aceleração gravitacional em três eixos, $\{x, y, z\}$. Considerando que o dispositivo móvel esteja no interior do veículo, as vibrações geradas pelo seu deslocamento serão refletidas no aparelho. Em média, os acelerômetros presentes nos dispositivos são capazes de realizar 160 leituras por segundo, o que pode gerar uma grande quantidade de dados em uma leitura contínua. Cada 50 metros percorridos representa um **trecho**. As características são extraídas para cada trecho, salvando na memória do dispositivo.

O sensor de geolocalização fornece a latitude, longitude, altura em relação ao nível do mar e velocidade corrente do dispositivo. O *Streetcheck* salva as coordenadas do veículo no início e ao fim do trecho de 50 metros. O conjunto de características extraídos no pré-processamento são descritos na Seção 3.3.

Durante o processo de coleta de dados, o usuário pode classificar um segmento de via como “BOM” ou “RUIM” gerando um rótulo para aquela amostra. No entanto, no contexto de classificação de vias, a concepção do que é um trecho ruim ou bom pode

variar entre os usuários. Nas situações em que uma mesma amostra de via é classificado de maneira adversa, o rótulo atribuído pela maioria é o escolhido. Baseando-se nisso, o sistema mantém um registro com todas as características dos trechos percorridos por cada usuário, com suas respectivas classificações – se houver. Dessa forma, os algoritmos de aprendizagem são capazes de fazer a classificação de novos trechos com base nesse conjunto de dados já rotulados.

A Figura 2 apresenta a interface gráfica desenvolvida para permitir que os usuários rotulem os trechos como “BOM” ou “RUIM”, além de fornecerem outros atributos sobre a viagem.

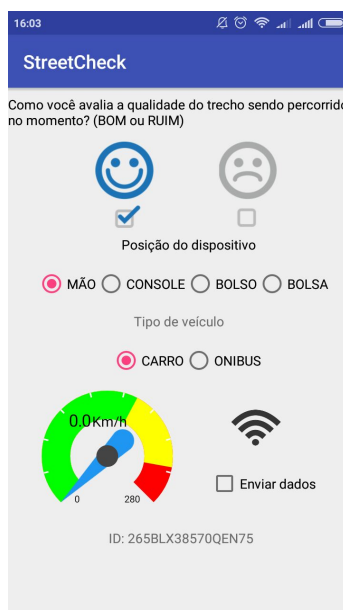


Figura 2. Interface do StreetCheck

3.3. Sistema de Classificação de vias

O sistema divide o conjunto de dados capturados em duas partes. A primeira é composta pelas informações de trechos rotulados pelos usuários que, neste caso, são utilizados como dados de treinamento. O segundo conjunto de dados é composto por todos os registros que os usuários não classificaram.

A partir desses dois conjuntos de dados, foram extraídos um conjunto de características para os algoritmos de classificação. Para validar o *Streetcheck* foram utilizados os seguintes algoritmos de aprendizado não supervisionados:

Random Forest : Constrói uma série de árvores de decisão e faz a classificação dos dados para cada uma das árvores geradas. Ao final, a classificação é feita com base na regra do voto majoritário, na qual, para cada amostra é atribuído o rótulo mais frequente dentre os classificadores.

SVM : Gera um hiperplano multidimensional que melhor separa o conjunto de dados em dois grupos. Para isso, é criado um modelo a partir de um conjunto de treinamento que contenha dados classificados.

KNN : Classifica as amostras com base na classe das amostras vizinhas. Para cada uma das amostras a serem classificadas, é calculada a distância entre as K amostras de treinamento mais próximas. A classe mais predominante entre os vizinhos é atribuída a amostra.

J48 : Gera uma árvore de classificação a partir de um conjunto de dados de treinamento. Cada nó realiza uma análise da relevância dos atributos e são selecionados aqueles que melhor dividem o conjunto de dados.

Nos algoritmos de classificação, os dados são agrupados de acordo com as semelhanças encontradas em suas características. O objetivo de cada característica é proporcionar uma visão distinta sobre cada aspecto das amostras e assim guiar o algoritmo de classificação no processo de predição.

O vetor $\{x, y, z\}$ do acelerômetro não representa uma boa característica para os algoritmos de aprendizagem. Por este motivo, foram propostas transformações sobre este vetor para refletir a qualidade da via. As características foram extraídas em três momentos: (i) instantaneamente, juntamente com a leitura do vetor do acelerômetro; (ii) pré-processamento, dispositivo sumariza a característica a cada 50 metros percorridos; e, por fim, (iii) pós-processamento, características geradas após transmissão dos dados. A Tabela 1 sumariza as características extraídas para avaliar a qualidade da via. Em seguida, cada uma destas características é detalhada.

Tabela 1. Conjunto de características

Característica	Momento de geração
Vetor Resultante	Instantâneo
Velocidade	Instantâneo
Soma	Pré-processamento, no aplicativo
Desvio padrão	Pré-processamento, no aplicativo
Suavidade	Pré-processamento, no aplicativo
Distorção	Pós-processamento, no servidor
Quartis	Pós-processamento, no servidor

3.3.1. Vetor Resultante - R

A partir do vetor $\{x, y, z\}$, fornecido pelo acelerômetro, o vetor resultante R é composto pela soma vetorial dos três eixos. A Figura 3(a) demonstra a magnitude do vetor resultante comparado aos três eixos do acelerômetro. Conforme demonstrado na Fig. 3(b), a magnitude do vetor resultante apresenta valores distintos para trechos considerados bons e ruins, representando as vibrações sofridas pelo veículo. É importante ressaltar que essa característica é apenas uma base para o cálculo de novas características, descritas a seguir.

3.3.2. Velocidade Média - VM

A velocidade do veículo se mostrou-se relevante quando analisada em conjunto com outras características. Por isto, a cada segundo é lido a velocidade do veículo e após cada

trecho (50 metros), calcula-se a velocidade média.

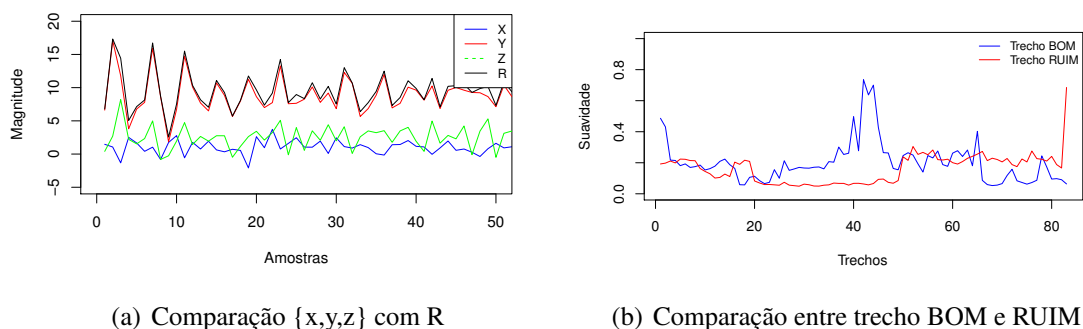


Figura 3. Vetor Resultante (R) calculado para cada amostra do acelerômetro.

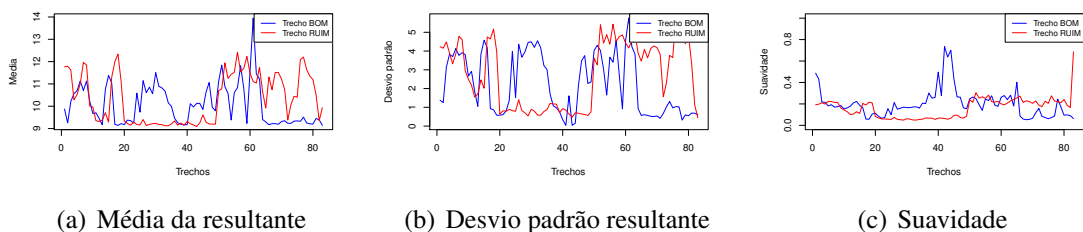


Figura 4. Comparação entre trechos bons e ruins de acordo com as características média, desvio padrão e suavidade.

3.3.3. Média do vetor resultante

O valor médio dos vetores resultantes em trechos percorridos pode representar as irregularidades da via. A Figura 4(a) ilustra a média dos vetores resultantes em uma série de trechos rotulados como BONS ou RUINS. Nota-se que os trechos ruins tendem a ter uma magnitude maior.

No entanto, o valor médio em um trecho bom pode ser maior do que em um trecho ruim. Isto ocorre devido a fatores como velocidade, posicionamento do dispositivo ou mesmo classificação inconsistente por parte do usuário. Logo, é necessário uma massa de dados consistente para aumentar a relevância desta característica.

3.3.4. Desvio padrão do vetor resultante - DP

O desvio padrão indica o quão as amostras estão espalhadas em relação à média. Deste modo, em pavimentos irregulares, o desvio padrão do vetor resultante (R) será superior. A Figura 4(b) ilustra o comportamento do desvio padrão em uma série de trechos bons e ruins. Assim como a média, o desvio padrão também é influenciado pela velocidade ou qualidade do rótulo atribuído pelo usuário. Porém, esta característica distingue melhor os trechos. Os picos na figura podem ser ocasionados por buracos, junções ou irregularidade abruptas na via.

3.3.5. Suavidade - SV

A característica definida como suavidade (SV) visa normalizar o desvio padrão do vetor resultante e a velocidade média do veículo em $[0, 1]$. Parte do princípio de que quanto mais regular e plana é a superfície, menor será o desvio padrão da amostra. A suavidade é definida pela Equação 1, sendo VM a velocidade média no trecho e DP o desvio padrão do vetor resultante.

$$SV = \frac{VM}{1+DP} \quad (1)$$

Quanto mais próximo SV for de 1, mais plano e regular é o trecho percorrido e quanto mais próximo de 0 maior o número de irregularidades. A Figura 4(c) ilustra o comportamento da característica suavidade. A curva dos trechos considerados bons são mais próximas de 1. No entanto, há interseção entre as curvas devido a imprecisão dos rótulos por parte dos usuários.

3.3.6. Distorção Média - DISTM

Distorção Média (DISTM) considera cada triângulo formado pelos eixos $\{x, y, z\}$ do acelerômetro como um plano. A partir disso, calcula o ângulo entre planos subsequentes. A soma dos ângulos de todos os planos encontrados no trecho é dividido por 360, obtendo-se o índice de distorção do trecho. Quanto menor este valor, mais plana é a via, com menos irregularidades.

3.3.7. Quartis - QX1, QX3, QY1, QY3, QZ1, QZ3

Para comparar os picos superiores e inferiores de cada eixo $\{x, y, z\}$ do acelerômetro, analisou-se o primeiro e o terceiro quartil de cada um. As Equações 2 e 3 calculam o primeiro e o terceiro quartil para cada eixo $E \in \{x, y, z\}$, respectivamente. Dessa maneira, QX1 e QX3 correspondem ao primeiro e terceiro quartis do eixo x . O mesmo princípio foi aplicado para os eixos y e z , gerando características independentes para cada eixo.

A Figura 5 ilustra a magnitude das amostras em um trecho em ordem crescente para cada eixo. Neste trecho, as categorias BOM e RUIM se diferem nos eixos no terceiro quartil (Q3) x e no primeiro quartil (Q1) do eixo y , enquanto o eixo z não apresenta grandes diferenças. Isto indica que a combinação destas características pode aumentar a taxa de acerto dos algoritmos de treinamento.

$$QE1 = \frac{\sum_1^{\frac{n}{4}} Ei}{\frac{n}{4}} \quad (2)$$

$$QE3 = \frac{\sum_{(\frac{n}{4}*3)}^n Ei}{\frac{n}{4}} \quad (3)$$

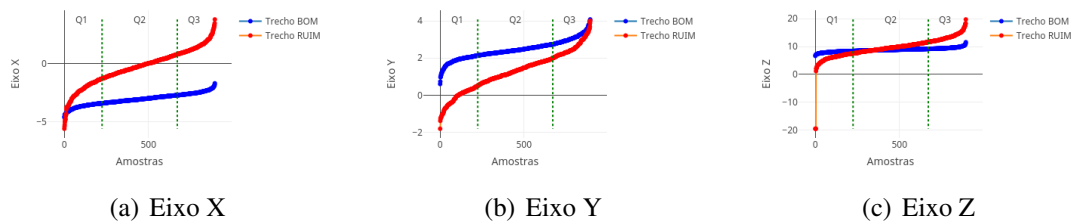


Figura 5. Comparação entre os quartis referentes aos eixos X, Y e Z

4. Classificando a qualidade de vias

4.1. Metodologia

Para validar o arcabouço *Streetcheck* e as características propostas, os testes foram divididos em duas fases: i) coleta de dados e, ii) classificação dos dados.

4.1.1. Coleta de dados

Na fase de coleta de dados, vinte (20) voluntários foram selecionados e instalaram o aplicativo em seus dispositivos móveis pessoais. Estes usuários eram responsáveis por garantir que o aplicativo estivesse em funcionamento durante trajetos em vias urbanas e quando possível, rotulá-los. A fase de coleta de dados durou três semanas.

Todos os voluntários foram orientados para apenas rotularem os dados quando fossem passageiros nos veículos. Quando o voluntário estivesse no papel de motorista, os trechos não são rotulados.

4.1.2. Classificação dos dados

Na segunda fase, foram analisadas se a quantidade de categorias da via (BOM ou RUIM) estava apropriada, a relevância de cada característica proposta e a taxa de acerto nos algoritmos supervisionados Random Forest, SVM, KNN e J48.

O *Streetcheck* permite ao usuário rotular a via de modo binário, isto é, BOA ou RUIM. Esta rotulação foi definida após analisar em testes preliminares o agrupamento em categorias dos dados coletados. Para isto, foi utilizado o algoritmo Kanopy [Whizbang et al. 2000] para verificar a curva de silhoueta [Rousseeuw 1987] e a melhor quantidade de categorias que expressam os dados.

Além disto, a presença de características irrelevantes ou redundantes em uma base de dados, pode afetar negativamente o desempenho de algoritmos classificadores. O problema de seleção de características pode ser definido como a busca por M características relevantes a partir de um conjunto de N atributos, de forma que $M \leq N$ [Liu and Setiono 1996].

Para definir a relevância de cada característica, apresentadas na Seção 3.3, foi utilizado o algoritmo *Relief* [Kira and Rendell 1992]. O *Relief* lida com dados de classe binária e produz como resultado um vetor $\delta(X)$, sendo X a característica analisada. Quanto

maior o valor obtido em $\delta(X)$, maior a relevância da característica. O *Streetcheck* utiliza a biblioteca GNU WEKA³ como implementação dos algoritmo Kanopy, *Relief* e todos os classificadores analisados.

4.2. Sumário dos dados coletados

A distância total percorrida e não classificada foi de 1218.34 km, sendo 1036.64 km em carros e 181.70 km em ônibus. Foram percorridos trechos entre as cidades de Belo Horizonte, João Monlevade, Ouro Preto, Ipatinga e Caratinga. No total, foram coletados 2,4GB de dados brutos pelos voluntários.

Do total de trechos percorridos e rotulados, que se tornam dados de treinamento, 63,72% foram de trechos considerados bons e 36,28% ruins. Alguns poucos trechos (0,6%) foram registrados mais de uma vez na base de treinamento com reclassificação, ou seja, quando a opinião muda de bom para ruim ou vice-versa.

5. Análise e Discussão dos Resultados

5.1. Quantidade de grupos representativos

Inicialmente, foi utilizado o algoritmo Kanopy [Whizbang et al. 2000] para identificar categorias distintas sobre a massa de dados. A Figura 6 exhibe o espalhamento dos dados quando considerado as características velocidade (eixo x) e suavidade (eixo y). Observa-se que o aumento da velocidade diminui o valor da suavidade. De fato, quanto maior a velocidade do veículo, irregularidade na via podem ser disfarçadas pela suspensão do mesmo. Importante ressaltar que a Figura 6 ilustra o agrupamento para apenas duas características.

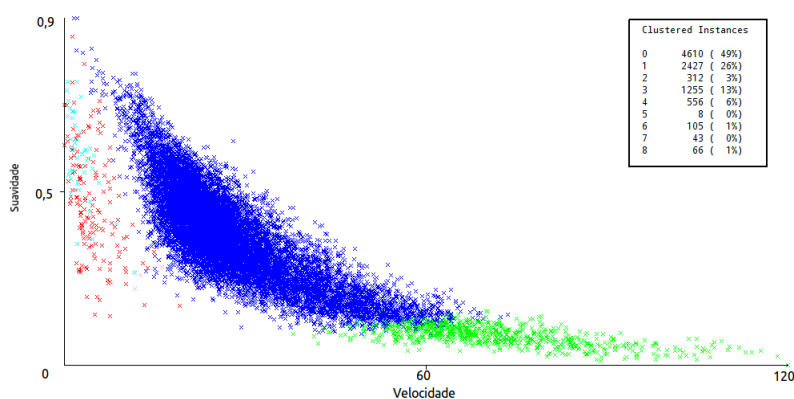


Figura 6. Clusterização usando algoritmo Kanopy

Para verificar o número de grupos que represente as características, foi utilizado a técnica de silhoueta, que avalia o quão bem ajustados os grupos (*clusters*) estão em relação aos dados. A Tabela 3 apresenta o fator silhoueta para cada quantidade de grupos (*clusters*) extraídas do dado não classificado.

Pode-se observar que, à medida que o número de categorias aumenta, a qualidade do fator de silhoueta decresce. Portanto, duas classes de dados são mais adequadas para o

³<https://www.cs.waikato.ac.nz/ml/weka/>

problema. Por este motivo, o *Streetcheck* classifica a qualidade das vias de modo binário (BOM ou RUIM).

Tabela 2. Fator de silhueta para clusters.

Clusters	2	3	4	5	6	8	10	15	20
Silhoueta	0.75	0.55	0.48	0.44	0.37	0.36	0.36	0.33	0.32

5.2. Comparação dos algoritmos

Em geral, nos problemas de classificação, é necessário definir qual algoritmo terá maior acurácia para classificar os dados. Uma abordagem simples, utilizada neste trabalho, consiste em estimar a acurácia de algoritmos de aprendizado supervisionado clássicos [Kotsiantis et al. 2007]. Foram utilizados 9.382 trechos rotulados (BOM ou RUIM) como base de treinamento e teste para os algoritmos KNN, Random Forest, J48 e o SVM.

Inicialmente, as características foram ranqueadas utilizando o algoritmo *ReliefF* [Houei 2005], que pontua a relevância de cada uma delas de acordo com sua capacidade de diferenciar as amostras. A Tabela 3 apresenta a relevância, de acordo com o algoritmo Relief, das onze características utilizadas.

Tabela 3. Lista de características e relevância.

Feature	VM	SV	QX3	QX1	DP	QZ1	QY1	DISTM	QY3	DISTM	QY3
Score ($\times 10^{-2}$)	3.079	1.328	1.228	1.202	1.032	0.949	0.916	0.882	0.835	0.825	0.795

Apesar de algumas características apresentarem baixa relevância, as mesmas combinadas às outras, incrementaram a acurácia dos classificadores. A Figura 7 apresenta a acurácia dos classificadores em relação ao número de características.

Utilizando apenas a velocidade média, os classificadores atingiram até 70% de acurácia. Os testes foram repetidos adicionando uma característica por vez por ordem de relevância (Tab. 3). Observou-se que a taxa de acerto do classificador se manteve a mesma ao utilizar apenas as oito características mais relevantes.

Enquanto os algoritmos SVM, KNN e J48 tem desempenho semelhante para quaisquer número de características, o algoritmo Random Forest se destaca e obteve a maior acurácia em todos os testes. A Tabela 4 apresenta a acurácia dos algoritmos utilizando as 11 características.

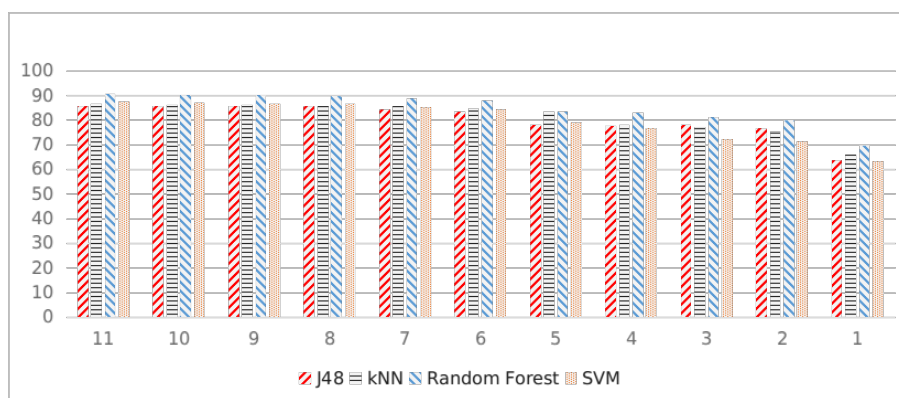


Figura 7. Acurácia dos algoritmos em relação ao número de características.

Tabela 4. Algoritmos utilizados e suas respectivas taxas de acerto

Algoritmo	Acurácia
Random Forest	90,64%
SVM	87,02%
Knn 5 vizinhos	86,05%
J48	85,71%

No problema tratado, a qualidade da base de testes depende dos rótulos informados pelos usuários à cada trecho. Em alguns trechos semelhantes, usuários inseriram classificações conflitantes. Além disso, foi observado que a velocidade gera ruídos em relação a percepção de impacto pelo acelerômetro. Para superar estes problemas, as características foram elaboradas de forma a acrescentar diferentes pontos de vista sobre os dados. Como resultado, não é apenas uma característica que classifica a amostra, mas todo o conjunto das características que é classificado.

Conforme descrito na Tabela 4, o algoritmo que obteve o melhor resultado foi o Random Forest por utilizar diferentes árvores de decisão em seu processo de classificação. Apenas uma árvore pode não ser suficiente para distinguir bem amostras com um número grande de atributos, sendo assim combinações diferentes são usadas em cada árvore e a classe mais votada é tida como a classe da amostra.

5.3. Visualização

O objetivo de todo sensoriamento participativo é que a colaboração individual de cada um, quando agregada, permita gerar informação para o coletivo. O *Streetcheck* permite ao usuário visualizar os mapas de classificação. A Figura 8(a) apresenta trechos no mapa que o usuário classificou, no qual os pontos azuis representam os trechos classificados como bons e os vermelhos como ruins.

No entanto, os dados coletados coletivamente pela comunidade de usuários permite classificar outros trechos. A Figura 8(b) apresenta os trechos classificados coletivamente. Desta forma, ao consultar o mapa, um usuário pode conhecer a qualidade de vias que ainda não visitou.

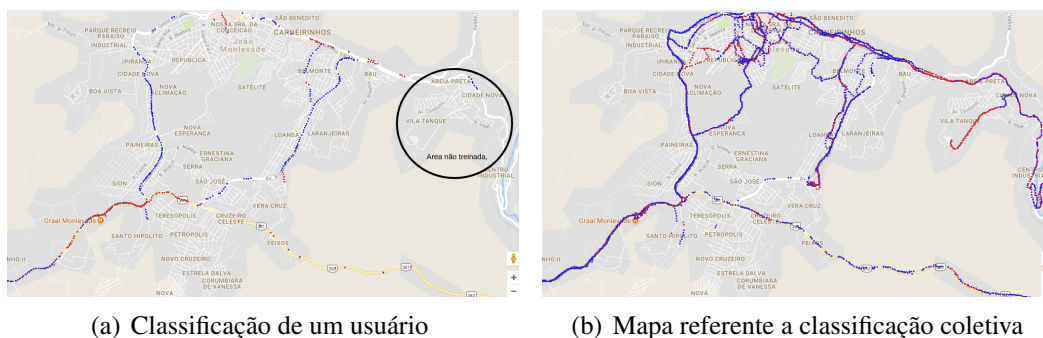


Figura 8. Mapas de classificação de vias

6. Conclusão

Este trabalho apresentou o StreetCheck, um sistema participativo, que permite aos usuários rotularem as vias em que transitam. Embora a opinião do usuário sobre a qualidade da

via seja qualitativa, diferentes usuários podem ter percepção diferente sobre a qualidade de uma mesma via baseada no contexto em que estão inseridos.

No contexto deste trabalho, o grande desafio é a extração de características que permitam aos algoritmos de aprendizado diferenciar corretamente a classe dos dados. A quantidade de ruídos produzido na coleta destas informações é alta, uma vez que sistemas dessa natureza estão sujeitos a corretude dos sensores nos dispositivos móveis dos usuários. Além disso, nem todas as características escolhidas dizem respeito às particularidades da qualidade das vias percorridas. Um exemplo disso é a velocidade média do veículo, característica que exerce influência sobre todas as outras, visto que a trepidação sofrida por um veículo depende da velocidade do mesmo. Este aspecto reforça a necessidade de se extrair um conjunto de características adequado ao contexto urbano, visto que a acurácia dos algoritmos classificadores depende das características extraídas.

A própria subjetividade do usuário pode representar um ruído na base de dados. A opinião do usuário é um ativo importante mas não é essencial para que este seja capaz de obter informações referentes a qualidade das vias. Nestes casos, o classificador utiliza apenas os dados gerados pelo acelerômetro e para rotular os trechos, depende da classificação de outros usuários.

Neste trabalho, com o auxílio de voluntários, foi possível extrair um conjunto de características que serviu de entrada para cinco algoritmos clássicos de aprendizado supervisionado. Por meio da análise dos dados coletados, foi possível demonstrar que duas categorias são suficientes para o problema de classificação de vias (boas ou ruins). A característica capaz de melhor representar a qualidade da via foi a suavidade, que leva em consideração a velocidade do veículo e o desvio padrão das leituras do acelerômetro. Além disso, mostrou-se que o algoritmo *Random Forest* obteve a maior taxa de acerto (90.64%) para classificar vias não rotuladas.

Como trabalhos futuros, pretende-se aumentar a massa de dados disponibilizando o aplicativo desenvolvido para o público geral. Deste modo, pode-se propor novas características que permitam classificar diferentes tipos de terrenos. Atualmente apenas o pavimento de asfalto é considerado, contudo, na realidade brasileira diversos tipos de terreno são comuns. Pretende-se também melhorar a *interface* do aplicativo para auxiliar os usuários no momento da escolha de rotas.

Referências

- Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., and Srivastava, M. (2006). Participatory sensing. In *In: Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications*.
- CNT (2016). Pesquisa cnt de rodovias.
- Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S., and Balakrishnan, H. (2008). The pothole patrol: using a mobile sensor network for road surface monitoring. In *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pages 29–39. ACM.
- González, L. C., Moreno, R., Escalante, H. J., Martínez, F., and Carlos, M. R. (2017). Learning roadway surface disruption patterns using the bag of words representation. *IEEE Transactions on Intelligent Transportation Systems*, PP(99):1–13.

- Houei, D. L. (2005). *Uso do ReliefF para Seleção de Atributos em Dados Multirrótulo*. PhD thesis.
- Kira, K. and Rendell, L. (1992). The feature selection problem: Traditional methods and a new algorithm. *AAAI-92 Proceedings*.
- Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160:3–24.
- Lima, L. C., Amorim, V. J. P., Pereira, I. M., Ribeiro, F. N., and Oliveira, R. A. R. (2016). Using Crowdsourcing Techniques and Mobile Devices for Asphaltic Pavement Quality Recognition. In *2016 VI Brazilian Symposium on Computing Systems Engineering (SBESC)*, pages 144–149. IEEE.
- Liu, H. and Setiono, R. (1996). Feature selection and classification—a probabilistic wrapper approach. *Proceedings of Ninth International Conference on Industrial and Engineering Applications of AI and ES*, pages 419–424.
- Mednis, A., Strazdins, G., Zviedris, R., Kanonirs, G., and Selavo, L. (2011). Real time pothole detection using Android smartphones with accelerometers. In *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*, pages 1–6. IEEE.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C):53–65.
- Silva, T. H., de Melo, P. O. V., Neto, J. B., IJT, A., Ribeiro, C. S. d. S., Mota, V. F., da Cunha, F. D., Ferreira, A. P., Almeida, J. M., and Loureiro, A. A. (2016). *Pervasive Computing: Next Generation Platforms for Intelligent Data Collection*, chapter Users in the Urban Sensing Process: Challenges and Research Opportunities. Morgan Kaufmann - Elsevier.
- Whizbang, A. M., Nigam, K., and Ungar, L. H. (2000). Efficient clustering of high-dimensional data sets with application to reference matching.