

Como inferir o meio de transporte das pessoas em cenários urbanos a partir de smartphones

Carlos Alvaro de M. S. Quintella, Carlos Alberto V. Campos, Leila C. V. Andrade

Programa de Pós-Graduação em Informática – PPGI

Universidade Federal do Estado do Rio de Janeiro (UNIRIO) – Rio de Janeiro – Brasil

caquintella@gmail.com, leila@uniriotec.br, beto@uniriotec.br

Abstract. *This paper presents a mechanism to classify the transportation mode used by people in urban scenarios through movement traces collected by smartphone's own (mechanism user). In addition, a performance evaluation of the classifiers is conducted in order to provide valuable information for use by systems context-aware, especially through middleware with cloud based processing.*

Resumo. *Este trabalho apresenta um mecanismo para classificar o meio de transporte utilizado pelas pessoas em cenários urbanos por meio de traces (registros) de movimento coletados por smartphone da própria pessoa (usuário do mecanismo). Além disso, uma análise do desempenho dos classificadores do modo de transporte é realizada com o objetivo de fornecer informações valiosas para sua utilização por sistemas cientes de contexto, em especial através de middleware com processamento na nuvem.*

1. Introdução

Foi através dos *smartphones* que a visão da computação ubíqua introduzida nos anos 90 por Mark Weiser, pesquisador da Xerox, está sendo efetivamente realizada de acordo com [Baudalf 2007]. Assim, hoje uma boa parte da geração e do consumo de conteúdo digital passou a ocorrer com o usuário em movimento, dado tanto pela ampla disponibilidade das redes sem fio quanto dos dispositivos móveis.

Nos últimos anos, diversos estudos se voltaram para entender as informações produzidas por usuários de *smartphones* durante suas utilizações diárias. Destacam-se os estudos: [Zheng 2008a, Zheng 2009, Zheng 2010, Reddy 2010, Stenneth 2011, Shin 2015, Shafique 2016, Zhoua 2016]. Através destes estudos é possível constatar que a capacidade de identificar o contexto do meio de transporte das pessoas nas cidades é importante para diferentes tipos de aplicação, como: sistemas de *e-health*, marketing, cidades inteligentes, gerenciamento de trânsito, auditoria e de segurança em seus mais variados espectros. Cabe ressaltar que a detecção desses meios de transporte em tempo de utilização requer um alto nível de acurácia, sob pena de que os usuários percam a confiança na aplicação e, por consequência, deixem de usá-la [Reddy 2008].

Dentro desse contexto, o presente trabalho tem o objetivo de apresentar uma técnica aplicada para inferir o meio de transporte das pessoas em ambientes urbanos a partir de *traces* (registros de posicionamento) de mobilidade de *smartphones*, gerados pelo deslocamento dos usuários portadores destes *smartphones*. A partir do presente trabalho têm-se as seguintes contribuições: (i) um modelo para representação de trajetórias a partir de registros de movimentação de *smartphones*; (ii) um mecanismo

para segmentar trajetórias em agregados temporais, (iii) uma proposta para implementação de classificadores de *traces* de *smartphones* compatível com o uso de middleware de sistemas cientes de contexto, possivelmente baseados em uma nuvem computacional, (iv) uma coleta de dados de mobilidade humana em um cenário real (regiões das cidades: Rio de Janeiro e Niterói) e, (v) uma lista dos melhores classificadores para o meio de transporte de acordo com os métodos apresentados.

Este trabalho se estrutura da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. Na Seção 3 são apresentados os principais métodos empregados a coleta de dados em um cenário real. A Seção 4 apresenta as atividades empregadas para a descoberta de conhecimento em banco de dados. A Seção 5 apresenta os resultados obtidos. Por fim, a Seção 6 apresenta a conclusão do artigo e os trabalhos futuros.

2. Trabalhos relacionados

No conjunto de trabalhos, [Zheng 2008a, Zheng 2008b, Zheng 2009], os autores abordam processos para classificar a movimentação dos usuários baseado na coleta de dados de movimentação, com uso do GPS embarcado em telefones celulares. Foi proposto um mecanismo para segmentar as trajetórias em etapas de movimentação, ou segmentos de trajetória de acordo com pontos de mudança. Os autores levaram em consideração que entre a troca de meios de transportes deve haver uma pausa, mesmo que discreta. Com os pontos de mudança detectados, os autores descobriram que quando ocorre a mudança de meio de transporte, deve haver um segmento que é do tipo andando, como apresentado em [Zheng 2010]. Apresentaram também, um mecanismo de detecção de meio de transporte, que se utiliza de limiares de distância e tempo para classificar uma movimentação, em tipos distintos: andando e não andando. Nos segmentos de trajetória, os atributos analisados foram: velocidade, aceleração, mudança de direção e paradas. Assim, para inferir o meio de transporte utilizado pelos usuários, foi usado um processo sequencial baseado em quatro etapas: segmentação, extração de características, inferência e pós-processamento.

Cabe ressaltar que inicialmente a aplicação do estudo não foi voltada para a identificação dos deslocamentos em tempo real, mas sim para os deslocamentos que já haviam sido concluídos. Entretanto, como explicado em [Zheng 2010], os autores além de apresentarem em maiores detalhes o mecanismo de classificação baseado em pontos de mudança, também introduzem um modo de inferência, funcionando próximo ao tempo real. Este é baseado no treinamento com dados históricos, que são gerados pelo método apresentado no trabalho anterior [Zheng 2008b], possibilitando assim inferir os meios de transporte ainda em andamento. Foram utilizados *logs* (registros) de GPS, portanto, sem o uso da técnica de fusão de sensores. A utilização do mecanismo de identificação de pontos de mudança caracteriza um sistema de informações geográficas, no qual os próprios usuários são usados para alimentar a base de dados. O uso dos pontos de mudança caracteriza uma abordagem composta de duas etapas, nas quais são empregadas técnicas para melhorar os resultados da classificação obtida inicialmente.

No trabalho [Reddy 2008] apresenta-se uma técnica para identificação de meio de transporte através de *smartphones* com uso de GPS e do acelerômetro. O foco adotado foi a análise da variação de velocidade, pelo GPS, com o uso do acelerômetro de forma complementar, permitindo identificar com um alto grau de precisão os deslocamentos (ente os modos: parado, andando, motorizado e bicicleta). O uso da

combinação dos sensores mostrou um aumento de 10% na precisão, em contraste com a utilização somente do GPS. A taxa de amostragem utilizada foi 1Hz. O tempo para inferência inicial do meio de transporte foi de apenas 2 segundos transcorrido da coleta de dados. Neste trabalho, os autores analisaram o desempenho de diferentes classificadores: k-Nearest Neighbor, Decision-tree, Naive Bayes, Support Vector Machines, Continuous Hidden Markov Models (CH-MM) e um classificador em dois estágios compostos de Decision Tree associado à Discrete Hidden Markov Models - onde o segundo componente é aplicado como um filtro para eliminar os “ruídos” relativos a fronteiras de trocas de meio de transporte.

No trabalho [Stenneth 2010], os autores apresentam uma abordagem para detecção de meios de transporte utilizando informações de posicionamento de *smartphones* e informações geográficas diversas, como: pontos de ônibus, de linhas ferroviárias e informação do posicionamento dos ônibus em tempo real. Diferentemente dos estudos apresentados anteriormente, neste estudo a taxa de amostragem para o posicionamento utilizada foi de 30 segundos. Para a detecção do meio de transporte próximo ao tempo real, os autores utilizaram uma janela móvel de 1 minuto (equivalente a dois registros de posicionamento), onde utilizam os seguintes atributos: velocidade média, aceleração média, média de mudança de direção.

No trabalho [Shin 2015] os autores utilizaram apenas o acelerômetro dos *smartphones* para detectar o meio de transporte utilizado pelo usuário. Sendo que também se utilizaram dos conceitos de segmentação de trajetória e da separação entre segmentos de meios de transportes distintos por períodos de caminhada, de [Zheng 2008a]. A frequência de leitura utilizada foi 5s para computar as médias de aceleração, com as quais foi possível classificar os segmentos da trajetória com um algoritmo de quatro etapas. A coleta de dados foi feita em Zurique na Suíça, com 495 amostras de 30 usuários, os meios de transportes detectados foram: trem/bonde, carro e ônibus.

Com base nesses trabalhos, nas próximas seções do presente artigo, etapas foram realizadas para explorar o problema da classificação dos *traces* e obter informações para subsidiar a proposta de um mecanismo de classificação de *traces* que possa melhorar os resultados aqui apresentados.

3. Métodos utilizados e a coleta de dados realizada

Nesta seção serão apresentados os métodos empregados para a classificação do *traces* de movimentação e a descrição de uma coleta de dados reais.

Para selecionar a plataforma a ser utilizada, primeiramente foi necessário analisar a representatividade de mercado de cada plataforma, seguida de uma análise detalhada dos aspectos tecnológicos: disponibilidade de ferramentas de desenvolvimento, sensores, APIs, documentação e suporte. Foram avaliadas as plataformas Android, Windows Phone 7 e Apple iPhone. Sendo escolhida a plataforma iPhone da Apple por ser a que melhor atendeu o conjunto dos critérios avaliados. Cabe ressaltar que durante o processo de desenvolvimento e coleta de dados, a Google ainda não havia disponibilizado o Activity Recognition na plataforma Android.

Para localização foi utilizado o sensor de localização do iPhone, acessado pelo *framework CoreLocation* do iOS, apresentado em [Apple 2012]. Trata-se de uma fusão de sensores, que utiliza os dados do GPS, o *fingerprint* dos sinais de WiFi e da telefonia celular para estimar a localização com um grau ótimo de precisão. A seguir serão

apresentadas seções referentes à representação da trajetória analisada, as técnicas e métricas utilizadas e um detalhamento sobre a coleta de dados realizada.

3.1. Representação da trajetória

Muitas vezes, os dados brutos coletados tem pouco valor para mineração de dados, assim, é necessária uma primeira interpretação e extração de características (features) antes de usá-los. A associação de registros de posicionamento a trajetórias é uma prática comum em banco de dados de trajetórias. O processo de estimar uma trajetória através de *smartphones* consiste em um problema difícil, pois não há um sensor embarcado que seja preciso em todas as localidades, conforme apresentado por [Thiagarajan 2011].

No presente trabalho foi utilizada uma etapa de processamento aplicada após a coleta de dados para o agrupamento dos diversos registros de posicionamento em trajetórias. Tal etapa visa identificar e associar um grupo de registros de posicionamento à uma trajetória específica. Então, após esta etapa, cada trajetória passa ter seus períodos de deslocamentos e pausas identificados, com seus respectivos registros de posicionamento associados a um único determinado deslocamento ou uma única pausa. Esta etapa está fortemente baseada no trabalho de [Idrissov 2012], onde os autores apresentam um framework para agrupamento de registros de posicionamento.

3.2. Técnicas e métricas utilizadas para avaliação de classificadores

Entre as técnicas de avaliação de classificadores, este estudo usou a técnica de validação cruzada (cross-validation), conforme descrita em [Tan 2009], mais especificamente foi utilizado o método k-fold (com k=10). Esta técnica foi selecionada por ser um método bem conhecido e aceito, por ser usado nos trabalhos de [Zheng 2008a, Stenneth 2011] e por ser compatível com a ferramenta Weka (utilizada para esta etapa do estudo). Sua utilização se deu de forma análoga ao que foi apresentado nos dois referidos trabalhos.

Os termos *precision accuracy* e *recall accuracy*, apresentados em [Sokolova 2006], referem-se a métricas para medição de desempenho dos classificadores onde existe ocorrência de classes não balanceadas. Desta forma, *precision accuracy* determina a fração dos registros que realmente acabaram sendo positivos no grupo que o classificador declarou como positivo. Já o *recall accuracy* mede a fração do total de exemplos positivos corretamente classificados pelo classificador. Um *recall accuracy* alto reflete em que de todas as ocorrências para uma determinada classe, poucas foram as que deixaram de ser corretamente classificadas [Russel 2010].

3.3. Detalhamento sobre as ferramentas de desenvolvidas para a coleta de dados

Visando maior domínio sobre a qualidade da informação obtida, chegou-se a conclusão que coletar os próprios dados de mobilidade apresentariam oportunidades adicionais a pesquisa. Além de permitir a investigação de aspectos voltados às características de mobilidade local, levando em consideração [Noulas 2012].

Primeiramente, foi criado um framework, para suportar não só a coleta de dados, mas toda a estrutura de pesquisa. A partir deste framework foi implementada uma infraestrutura para subsidiar a coleta de dados. Também foi desenvolvida uma aplicação para iPhone, denominada CityTracks, cuja a interface é apresentada na Figura 1. Esta aplicação permite a coleta de dados de movimentação, além de outros dados relativos às pesquisas de sensoriamento participativo, como notas e fotos geo-localizadas.

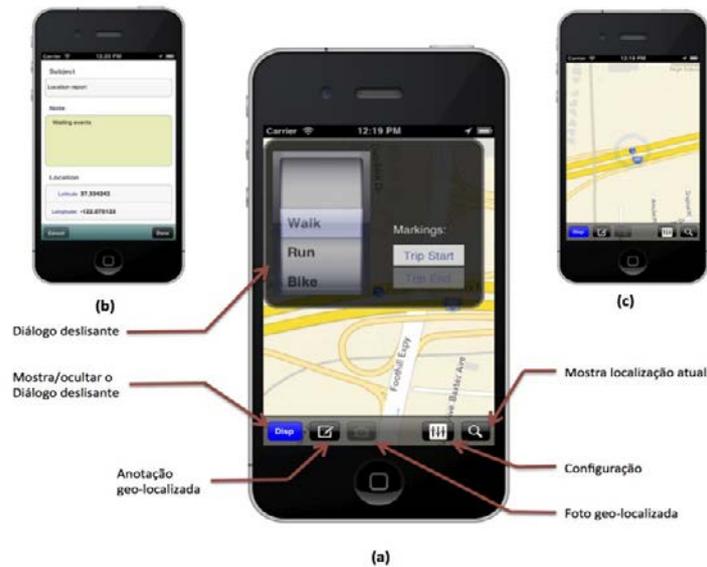


Figura 1. Interface da aplicação CityTracks. (a) tela principal, (b) tela entrada de notas, (c) tela visualização de posição.

Sobre a coleta de dados de movimento de pessoas foram utilizados nove voluntários, que disponibilizaram seus deslocamentos entre setembro/2012 e maio/2013 e, através dos quais, foram coletados mais de 880.000 registros de posicionamento. Os registros foram provenientes de 180 trajetórias desses voluntários, sendo que estas trajetórias referem-se aos seguintes meios de transporte: carro, motocicleta, ônibus, bicicleta, barca, a pé e metrô. A área geográfica abrangida pela coleta dos dados incluiu: (i) Zona Sul, Zona Norte e Centro da cidade do Rio de Janeiro e; (ii) Zona Sul, Zona Norte, Centro e Região Oceânica da cidade de Niterói. Os voluntários foram instruídos para que ao iniciar suas movimentações, executassem a aplicação *CityTracks* e informassem o início e o fim dos deslocamentos, junto com o meio de transporte utilizado. Foi através deste procedimento que viabilizou o uso de técnicas de classificação na etapa de mineração de dados. Cabe ressaltar que não foi possível trabalhar com os dados de trajetórias de metrô e barcas, devido à baixa qualidade e pequeno número de registros obtidos. Logo, as amostras relacionadas a este tipo de deslocamento foram descartadas. Assim, foram utilizados somente os dados de movimentação dos meios de transporte: ônibus, carro, moto, a pé e de bicicleta.

O formato do registro de localização coletado equivale aos dados disponibilizados pelo Corelocation framework do iOS6: *device_id*, *time_stamp*, *horizontal_accuracy*, *latitude*, *longitude*, *vertical_accuracy*, *altitude*, *speed*, *course*. Quanto a granularidade dos dados coletados, apesar de não ser possível controlar a frequência da geração dos registros no iPhone, em etapa posterior, quando a frequência ultrapassou 1 Hz, foi apenas mantido um dos registros do segundo e descartado os de menor *horizontal_accuracy*. Quando feita a comparação com os trabalhos relacionados, é possível notar as taxas de amostragem como: 30 segundos em [Stenneth 2011]; 1 segundo em [Reddy 2008] e, por fim, 2 segundos em [Zheng 2008a].

Ao longo do presente trabalho, nas etapas relacionadas à descoberta de conhecimento em banco de dados foi usado o método KDD descrito em [Fayyad 1996].

4. Análise de dados

Ao longo de todo processo apesar da utilização de algoritmos de classificação em trajetórias já concluídas, nenhum processamento ou suposição foi usado de forma que invalidasse a aplicação destas técnicas para deslocamentos em andamento. Com isso, espera-se que seu uso possa se dar para aplicações cientes de contexto e interativas.

4.1. Preparação e transformação dos dados

O processo de preparação dos dados foi feito através de scripts escritos no software *R*, sendo que um script para cada etapa. A primeira etapa correspondeu a preparação dos dados, ou pré-processamento. Nesta etapa foi feita a remoção dos registros imperfeitos, neste caso, onde o atributo *horizontal_accuracy* era maior do que 200 (metros), tendo como objetivo obter somente os registros com melhor precisão, excluindo *outliers*. A linha de corte de 200 m foi definida levando em consideração: (i) não descartar um número excessivo de registros, (ii) acomodar os diversos meios de transportes distintos, (iii) e ainda assim permitir o funcionamento em locais onde se tem mais baixa precisão de leitura, como embaixo de viadutos, áreas com arranha-céus e morros. Os registros foram então ordenados por dispositivo e tempo (*timestamp*).

A seguir, na etapa de transformação do dado foi feita a separação das trajetórias, orientada pelo framework de limpeza de trajetórias, apresentado em [Idrissov 2012], também se levando em consideração o trabalho [Zheng 2008a]. As trajetórias então foram derivadas em componentes de deslocamento ou de pausa. Sendo, que para isso foi necessário, não só a capacidade de detectar as pausas, como também o final de um deslocamento. No caso da detecção de pausa, esta foi feita pela aplicação de limiares de tempo e distância de pausa (0,4 metros e 90 segundos, respectivamente). Estes números foram obtidos pela análise visual dos deslocamentos e levaram em consideração condições de tráfego nas cidades do Rio de Janeiro e Niterói. Já para a trajetória utilizou-se o limiar de vinte minutos. Com isso, no caso deste limiar ser ultrapassado sem nenhuma ocorrência de registros de posicionamento, a trajetória corrente então passa a ser considerada encerrada, com os registros de posicionamento posteriores fazendo parte de uma nova trajetória. No caso de falta de alguns registros ao longo de uma trajetória, foi feita uma interpolação. Considerou-se a velocidade do valor anterior, considerando assim, que ocorreu um deslocamento uniforme entre o registro anterior e o registro corrente. Este tipo de procedimento produziu efeitos análogos a uma interpolação de pontos uniformemente distribuídos.

Com os segmentos de trajetória definidos foi possível analisar as informações de meio de transporte provenientes dos usuários, para então atribuir à classificação apresentada a um determinado segmento. Assim, cada registro de posicionamento, pode então ser associado a um deslocamento ou a uma pausa. Sendo que o deslocamento ou a pausa, por sua vez, são associados a uma determinada trajetória. Este procedimento é compatível ao apresentado no trabalho [Zheng 2011].

Dado que o objetivo deste trabalho é a classificação da movimentação que ainda está em andamento, foi utilizado um esquema de agregação temporal para agrupar uma série contígua de registros até que um determinado tempo seja atingido, gerando uma tupla de agregação com informações relativas ao período de tempo. Este agrupamento foi feito na última etapa de pré-processamento. Os segmentos temporais foram denominados *chunks*. Foram analisados três tamanhos de *chunks*: 60, 90 e 120 segundos. Sendo que os *chunks* de 90 segundos apresentam um melhor resultado quando se leva em consideração o ganho de precisão, inferência e o menor tempo para

início da classificação. Estes valores foram escolhidos tendo em vista não utilizar um valor pequeno demais, que poderia não incluir características de movimentação suficientemente ricas ou um valor grande demais, que poderia retardar o processo de identificação do meio de transporte.

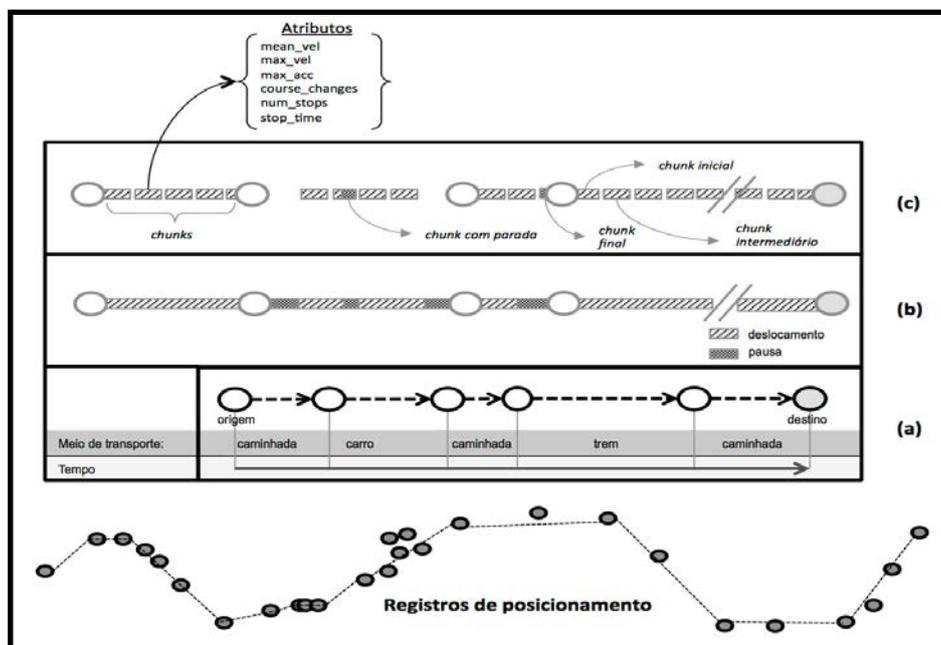


Figura 2. Representação da informação do movimento coletado passo a passo no processo de transformação e sumarização dos dados.

A partir dos *chunks* as informações de movimentação foram extraídas. Estas por sua vez podem ser traduzidas nos atributos: velocidade média; velocidade máxima, aceleração máxima, número de paradas e número de mudanças de direção, ilustrados na Figura 2. Para obtenção da velocidade foi utilizada a leitura de velocidade instantânea do GPS e a aceleração foi obtida através da computação da variação de velocidade ponto a ponto: $(v_{p2} - v_{p1})/t$, onde t é o intervalo de tempo e v_{p1} e v_{p2} , são respectivamente, velocidade reportada no ponto imediatamente anterior e a no ponto atual. Já quanto à mudança de curso (que é medido em graus radianos), este foi dividido por dez, onde somente a parte inteira da divisão foi considerada visando diminuir a sensibilidade a mudança de direção. Para o tempo de parada e número de paradas levou-se em conta a distância do deslocamento ponto a ponto dada pela fórmula de Haversine conforme descrita em [Cram 2016] e deslocamentos com velocidade menor que 0,4m/s foram considerados estacionários.

Na Figura 2 é mostrado como os registros de movimento são usados pela técnica proposta para inferir o meio de transporte. Na parte (a) desta figura é ilustrada a inferência já realizada dividindo o trajeto em diferentes meios de transporte. Na parte (b) a trajetória foi separada em porções de deslocamento e pausa. Na parte (c) podemos observar os *chunks* e seus atributos sendo identificados e representados.

4.2. Seleção de atributos e algoritmos

A etapa de seleção de atributos levou em conta tanto a análise visual da distribuição de frequência de cada um dos atributos e classes, quanto uma análise do ganho

informacional feita com o *Weka Experimenter*. Dessa forma, os seguintes atributos foram selecionados para utilização na etapa de mineração de dados: velocidade máxima, aceleração máxima, número de mudanças de direção.

A escolha de algoritmos foi feita com base nos trabalhos [Zheng 2010], [Reddy 2010], [Stenneth 2010] e [Wu 2008], onde o critério adotado para a seleção foi investigar tanto os algoritmos que apresentaram bons resultados em trabalhos anteriores, quanto outros que ainda não haviam sido verificados. Com isso, a lista de algoritmos analisados foi: Redes Bayesianas, Naive Bayes, SVM, Multilayer Perceptron, Decision Tree, Random Forest, Random Trees, K-Means, K-NN e Ada boost.

4.3. Avaliação dos classificadores

O processo de avaliação dos algoritmos de classificação foi feito com uso da ferramenta *Weka*. Para tanto, utilizou-se o método de *cross-validation 10-fold*. Iterações de análise sucessivas foram feitas, onde as etapas anteriores forneceram os fundamentos para as decisões de análise. Através desta abordagem foi possível realizar uma análise exaustiva e chegar aos melhores classificadores, inclusive para casos específicos de classificação.

Para treinamento foi utilizada uma aleatória amostragem equivalente a 20% dos dados, os demais 80% foram utilizados para a avaliação. Com isso, o número de *chunks* utilizados para o treinamento foi: carro=528; moto=144; andando=230; bicicleta=177, ônibus=54. Foi feita uma avaliação do impacto da amostra heterogênea pela aplicação de *oversampling* usando técnica de SMOTE, apresentado em [Tan 2009]. Assim, foi possível constatar a ausência de melhoria significativa para a classificação.

5. Resultados obtidos

A classificação dos *chunks*, segundo o meio de transporte, foi analisada com múltiplas abordagens e algoritmos: (i) classificar entre as classes andando, bicicleta, moto, carro e ônibus; (ii) classificar entre os meios motorizados e não motorizados; (iii) classificar apenas entre os não motorizados, ou seja: andando e bicicleta, (iv) classificar apenas entre os motorizados, ou seja, carro, moto, ônibus; (v) classificar entre andando, bicicleta e motorizado e, por fim; (vi) classificar em duas etapas, na primeira retirando algumas classes com alta taxa de classificação e, posteriormente, em uma segunda etapa fazendo a identificação das classes específicas. Com isso, foi possível identificar o conjunto de algoritmos com melhor resultado em cada caso, como descrito na Tabela 1.

Com base nos resultados obtidos na Tabela 1 foi possível chegar as seguintes conclusões: (i) - detectar um usuário andando não é tarefa difícil, pois diversos algoritmos conseguiram com precisão próxima a 100%; (ii) - detectar somente usuários andando e de bicicleta também não é difícil, pois os resultados ficaram acima de 90%; (iii) - separar as movimentações de usuários entre carro, moto e ônibus, é difícil (64%), pois os modos apresentaram perfis de movimentação muito parecidos; (iv) - separar usuários motorizados dos andando e de bicicleta, apresenta bons resultados (acima de 90%). Além disso, é possível utilizar a classificação em duas etapas, onde primeiramente extrai-se a possibilidade de alguns modos de movimentação. Depois, esta informação é usada de forma estocástica para a classificação na segunda etapa (vide Tabela 1). Os resultados se apresentaram mais promissores, subindo a precisão da classificação de 69% de acerto para 79% de acerto; quando se separando em três

classes: andando, bicicleta e motorizado na primeira etapa e, em seguida, aplicando o classificador para separar os motorizados se necessário.

Tabela 1. Resultado da análise dos meios de transporte por meio da acurácia de cada classificador utilizado para as oito classes de tipos de transporte

Classes	Abordagem	Classificador	acurácia
andando, bicicleta, moto, carro e ônibus	simples	DT	69%
andando, não-andando	simples	todos	> 98%
andando, bicicleta, motorizado		J.48	90%
motorizado, não motorizado	simples	SMO	92%
andando, bicicleta	simples	DT ou BN	94%
moto, carro, ônibus	simples	DT	64%
1) andando; bicicleta, motorizado; 2) carro, moto, ônibus	em duas etapas	J.48 e DT	79%
1) motorizado, não motorizado; 2a) andando, bicicleta; ou 2b) moto,	em duas etapas	SMO e DT	73%*

BN: BayesNet; DT: DecisionTable

6. Conclusão e trabalhos futuros

Foi apresentado um mecanismo para transformar dados brutos provenientes de traces de *smartphones*, em informação para subsidiar a extração do meio de transporte. Este mecanismo foi proposto com base na aplicação de aprendizado de máquina, e com base nos resultados de um processo de descoberta de conhecimento em banco de dados. Processo este que culminou com uma análise detalhada da performance de diversos classificadores; disponibilizando importantes informações para subsidiar a construção de um sistema de classificação da movimentação dos usuários. Tal mecanismo não necessita de etapas adicionais de pós-processamento e apresenta resultado superior a outras soluções abordadas entre os trabalhos relacionados, quando se compara as que somente utilizam o sensor de localização para fazer a inferência do meio de transporte.

Em relação aos trabalhos relacionados, é possível indicar um aumento próximo a 5% na precisão da detecção do meio de transporte sem o uso de sensores ou informações adicionais. Embora o conjunto de meios de transporte analisados e classes de inferência sejam diferentes, o presente trabalho apresenta o mesmo número que foi apresentado no trabalho com melhor resultado [Reddy 2008]. Cabe ressaltar ainda que o mecanismo aqui proposto utilizou dados para treinamento uma região geográfica específica, como em [Reddy 2008]. No caso de utilização do mecanismo aqui proposto com o uso de *middleware* para sistemas cientes de contexto baseados em nuvem, existe a possibilidade de utilizar o sistema de classificação apenas de acordo com o meio de interesse de cada aplicação. Com isso, a aplicação ciente de contexto poderia receber (através de uma assinatura) informações, somente para os meios de transporte de seu interesse. Podendo-se assim, potencializar o aumento da qualidade da identificação pela exclusão um ou mais meios de transporte; como por exemplo, uma aplicação de *e-health* poderia estar apenas interessada em diferenciar movimentações: andando, bicicleta e motorizadas. Outro exemplo seria uma aplicação de monitoração de pegadas de carbono, somente interessada nas movimentações motorizadas [Shin 2015].

Para trabalhos futuros existe a possibilidade de aumentar a precisão da inferência com o uso dos dados históricos de cada usuário ou analisar a possibilidade de aumentar a precisão à medida que o deslocamento vai acontecendo, ou seja, baseado nas medições anteriores. Além disso, existe também a possibilidade de se efetuar uma análise para bases de dados obtidas em diferentes cidades e comparar os resultados.

7. Referências

- Apple Inc., Location Manager Class Reference, disponível em: https://developer.apple.com/library/ios/#documentation/CoreLocation/Reference/CLLocationManager_Class/CLLocationManager/CLLocationManager.html#//apple_ref/doc/uid/TP40007125, acessado em 20/04/2016.
- Cram.org, Spherical trigonometry for geographic applications, página 16, disponível em: <https://cran.r-project.org/web/packages/geosphere/geosphere.pdf>, acessado em 01/06/2016.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., “The KDD process for extracting useful knowledge from volumes of data”, *Communications of the ACM*, v. 39, n. 11, 1996.
- Idrissov, A., Nascimento, M., Trajectory Cleaning Framework, Mobile Data Challenge (by Nokia) Workshop, 2012.
- Noulas, A., Scellato, S., Lambiotte, R., et. al., “A tale of many cities: universal patterns in human urban mobility”, *PLoS one*, Public Library of Science, v. 7, n. 5, 2012.
- Reddy, S., Burke, J. Estrin D., Et al. “Determining Transportation Mode On Mobile Phones”, *IEEE ISWC*, page 25-28, 2008.
- Reddy, S., Mun, M., Burke, J., et al., “Using mobile phones to determine transportation modes,” *ACM Transactions on Sensor Networks*, v. 6, n. 2, p. 13, 2010.
- Russel, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edition, New Jersey - EUA, Prentice Hall, 2010.
- Shafique, M. A., Hato E., “Travel Mode Detection with Varying Smartphone Data Collection Frequencies”, *Sensors*, 16(5), MDPI, p. 1-24, 2016.
- Shin, D., Aliaga, D., Tunçer, B., et al., “Urban sensing: Using smartphones for transportation mode classification”, *Computers, Environment and Urban Systems*, v. 53, p. 76–86, 2015.
- Sokolova, M., Japkowicz, N., Szpakowicz, S., “Beyond Accuracy, F-score and ROC: A Family of Discriminant Measures for Performance Evaluation”, *AAI 2006*, p. 1015-1021, Springer.
- Stenneth, L., Wolfson, O., Yu, P., et al., “Transportation mode detection using mobile phones and GIS information”, *19th ACM SIGSPATIAL*, p. 54-63, 2011.
- Tan, P., Steinbach, M., Kumar, V., *Introdução ao Data Mining, Mineração de Dados*, Rio de Janeiro - Brasil, Editora Ciência Moderna Ltda., 2009.
- Thiagarajan, A. “Probabilistic Models for Mobile Phone Trajectory Estimation”, Tese de doutorado, Massachusetts Institute of Technology, Setembro de 2011.
- Wu, X., Kumar, V., Quinlan, J. R., et al., “Top 10 algorithms in data mining”, *Knowledge and Information Systems*, 14(1), 1-37, 2008.
- Zheng, Y., Liu, L., Wang, L., et al., “Learning transportation mode from raw GPS data for geographic applications on the web”, *17th ACM WWW*, p. 247, 2008.
- Zheng, Y., Li, Q., Chen, Y., Xie, X. and Ma, W., “Understanding mobility based on GPS data”, *Proceedings of the 10th ACM Ubicomp*, p. 312-321, 2008.
- Zheng, Y., Zhang, L., Xie, X., and Ma, W., “Mining interesting locations and travel sequences from GPS trajectories”, *Proc. of the 18th ACM WWW*, p. 791-800, 2009.
- Zheng, Y., Chen, Q., Xie, X., et al., “Understanding transportation modes based on GPS data for web applications,” *ACM Transactions on the Web*, v. 4, no. 1, p. 1-36, 2010.
- Zhoua, X., Yub, W., Sullivan W. “Making pervasive sensing possible: Effective travel mode sensing based on smartphones”, *Computers, Environment and Urban Systems*, v. 58, 2016.