

CALIBRAÇÃO AUTOMÁTICA DE PARÂMETROS FUNDAMENTAIS DE TRÁFEGO EM APROXIMAÇÕES DE INTERSEÇÕES URBANAS SEMAFORIZADAS

Fábio de Souza Pereira Borges

Adelaida Pallavicini Fonseca

Programa de Pós-Graduação em Transportes, Universidade de Brasília

Li Weigang

Programa de Pós-Graduação em Informática, Universidade de Brasília

RESUMO

A coleção e análise dos dados registrados nos sistemas de controle de tráfego e equipamentos de fiscalização eletrônica de veículos baseadas em *big data* constituem provavelmente uma das ferramentas de pesquisa mais importantes e amplamente usadas em transportes. Neste contexto, as técnicas de *data mining* têm proporcionado excelentes resultados na classificação e previsão dos regimes de fluxo livre e congestionado em rodovias, exclusivamente. Então, considerando que vias urbanas e rodovias possuem comportamentos diferentes, este trabalho propõe uma metodologia de calibração dos parâmetros fundamentais de tráfego, específica para aproximações de interseções urbanas semaforizadas, por meio da técnica de agrupamento *k-means* aplicada aos dados registrados por esses equipamentos instalados nas aproximações. Para validar o método, foram utilizados os dados coletados por 08 sensores dispostos em duas interseções entre os meses de dezembro de 2014 e junho de 2017. Os resultados dessa metodologia são consistentes e promissores, principalmente devido à sua aderência às condições usualmente aceitas no âmbito da engenharia de tráfego e ao aproveitamento da infraestrutura de equipamentos de fiscalização existente nas cidades brasileiras.

ABSTRACT

Collection and analysis of data recorded on traffic control systems and its electronic sensors based on big data techniques is probably one of the most important and widely used research tools in transportation. In this context, data mining techniques have provided excellent results in the classification and prediction of highway's free and congested traffic flow regimes, exclusively. Therefore, considering that urban roads and highways present different behaviors, this work proposes a calibration method of fundamental traffic parameters regarding vehicles approaching urban road intersections with traffic lights, by means of the k-means clustering technique applied to the data recorded by traffic control electronic sensors installed near intersections. To validate the method, it was used the data collected by eight sensors located at two intersections between the months of December 2014 and June 2017. The results of this methodology are consistent and promising, mainly due to its adherence to the conditions usually accepted in the scope of traffic engineering and to the use of the traffic control electronic sensors infrastructure in Brazilian cities.

1 INTRODUÇÃO

Dunkel *et al.* (2011) ressaltam que muitos sistemas de controle de tráfego nos dias atuais são baseados em equipamentos de fiscalização eletrônica de veículos, que registram um volume muito grande de eventos individuais bem detalhados. Na mesma linha, Zeng (2015) frisa que, com o rápido desenvolvimento das cidades, a demanda por transporte motorizado criou a necessidade de monitoramento contínuo do tráfego, o que tem propiciado a instalação uma vasta gama de equipamentos de fiscalização nas redes viárias, gerando um espectro de dados cada vez mais dinâmico, extenso e diversificado. Essa estratégia trouxe um grande desafio aos analistas de trânsito que se dá pelo armazenamento, processamento e gerenciamento dessa enorme quantidade de dados (*Big Data*) produzidos por esses sistemas de controle. A exploração desse conjunto de dados, pode ser visto tanto como uma oportunidade quanto um desafio, especialmente para o gerenciamento e a governança de dados de transportes (Karlaftis e Vlahogianni, 2011; ITF, 2015).

A análise dos dados é importante em um sistema de controle de tráfego para garantir ações de controle que respondam rapidamente às mudanças detectadas no tráfego, o que passa obrigatoriamente pela previsão dos fluxos de tráfego. Tan *et al.* (2009) afirmam que a previsão de fluxos de tráfego é uma parte essencial do planejamento de transportes, do controle de

tráfego e dos sistemas de transportes inteligentes (ITS). No que diz respeito aos sistemas de controle de tráfego urbano, para Chen e Sun (2016), o desempenho dos sistemas adaptativos de controle de tráfego depende fortemente dos algoritmos de previsão dos fluxos de tráfego e de otimização dos semáforos e a redução do atraso dos veículos é essencial para a gestão do tráfego urbano, pois, de acordo com Guan e Yang (2002), tais sistemas podem coordenar a performance dos veículos nas redes viárias e minimizar o atraso dos veículos tanto quanto possível, mitigando os problemas causados pelo aumento da quantidade de veículos e do crescimento urbano acelerado.

Então, é importante que um modelo de previsão associado aos sistemas de controle de tráfego seja capaz de identificar quando o tráfego passa do regime livre para o regime congestionado e, de acordo com Sun e Zhou (2005), a segmentação dos dados coletados em fluxo livre e congestionado é realizada de forma exógena com base no melhor palpite do pesquisador e não existe um procedimento universalmente recomendado para determinar a ocupação crítica que define essa segmentação em um conjunto de dados de tráfego (Kianfar e Edara, 2013). Dervisoglu *et al.* (2009), Jiang e Huang (2009) e Li e Zhang (2011) sugeriram estruturas para calibração dos parâmetros fundamentais de tráfego utilizando os dados coletados por equipamentos de fiscalização eletrônica de veículos instalados em seções rodoviárias livres, em especial para produzir os diagramas Ocupação x Fluxo, Velocidade x Densidade, Densidade x Fluxo, respectivamente. Ainda no contexto do tráfego rodoviário, Kianfar e Edara (2013) pesquisaram uma estrutura para classificar o fluxo de tráfego em regimes de fluxo livre ou congestionado. A literatura revisada até a presente data não revelou ainda trabalhos dedicados às vias urbanas no contexto em epígrafe mediante o emprego de técnicas de *data mining*.

Assim, considerando a contextualização acima, o foco deste artigo são as vias urbanas e o seu objetivo principal é a proposição de uma metodologia de calibração automática dos parâmetros fundamentais de tráfego mediante o emprego de *data mining* em dados coletados por equipamentos de fiscalização eletrônica veicular instalados nas aproximações de interseções semaforizadas, aproveitando à grande disponibilidade desses equipamentos em cidades brasileiras.

De uma maneira geral, este trabalho está estruturado em 6 partes, sendo a primeira esta introdução. Na sequência, uma breve revisão da literatura apresenta os principais trabalhos que contribuíram para o desenvolvimento do tema. Depois, o item *Data Mining* introduz os conceitos fundamentais da mineração de dados, bem como as técnicas empregadas na construção da metodologia proposta, que é apresentada no quarto item da sequência deste artigo. Por fim, os resultados obtidos e as observações e discussões pertinentes a eles são abordados respectivamente nos itens 5 e 6 deste artigo.

2 REVISÃO DA LITERATURA

As estimativas dos parâmetros de tráfego dependem dos conceitos empregados, da aplicação dos métodos de determinação desses parâmetros e da aderência de modelos e métodos tradicionais ao fenômeno (Luna, 2003). Neste contexto, Kianfar e Edara (2013) indicam que um procedimento que não dependa completamente do julgamento dos pesquisadores pode permitir mais consistência para os resultados do estudo ou do modelo. Em um dos trabalhos mais citados sobre esse tema, Dervisoglu *et al.* (2009) calibraram um modelo de transmissão celular para uma seção rodoviária, utilizando um modelo de regressão linear ajustada aos dados de ocupação e fluxo somente com velocidades superiores a 88 km/h para criar o ajuste de fluxo

livre. Depois, eles empregaram a definição de capacidade do Manual de Capacidade Rodoviária (HCM, 2000) para identificar graficamente a capacidade dos segmentos das vias estudadas, tomando o maior fluxo registrado como a capacidade máxima da via. E comentam que na interseção da projeção horizontal desse fluxo sobre a reta de regressão linear ajustada aos dados selecionados é onde ocorre a ocupação crítica. Para concluir a estrutura, uma técnica de regressão quantílica aproximada é usada para criar o ajuste de ocupação e fluxo supercrítico.

Jiang e Huang (2009) propuseram um novo algoritmo baseado em técnicas de *data mining* para calibrar os parâmetros fundamentais de tráfego. Esses pesquisadores desenvolveram uma técnica de agrupamento hierárquico baseado no método *k-means*, que seria uma combinação deste método com a classificação aglomerativa hierárquica. De acordo com esse trabalho, o novo algoritmo foi capaz de reduzir os erros inerentes ao estágio inicial da classificação hierárquica aglomerativa, que resultou em um agrupamento com desempenho melhor. Como variáveis, foram utilizadas densidades e fluxos capturados por sensores de um segmento rodoviário de Pequim a fim de melhorar a precisão da calibração paramétrica. Os resultados dessa abordagem mostraram que o desempenho do algoritmo proposto foi melhor do que aqueles testados pelas soluções tradicionais.

Li e Zhang (2011) apresentaram um método para criar e calibrar um diagrama densidade versus fluxo. Primeiro, com base nas características das flutuações das séries temporais, os dados de tráfego são particionados em fluxo livre e fluxo congestionado. Depois, uma técnica de princípio mínimo é usada para identificar estados de equilíbrio seguidos pela aplicação de uma técnica mista de otimização de números inteiros para criar ajustes lineares de densidade e fluxo por partes. O objetivo do procedimento de otimização foi obter ajustes com desvio absoluto mínimo. A metodologia sugerida foi aplicada em três rodovias localizadas na Califórnia e, segundo os autores, os seus resultados aderiram muito bem aos princípios da engenharia de tráfego.

Ainda explorando as estruturas de calibração de parâmetros associados à densidade e ao fluxo de tráfego, Jiang *et al.* (2012) empregaram um volume maior de informações detectadas em uma rodovia para melhorar a precisão no processo de calibração de parâmetros usados em um modelo de simulação de tráfego. A mineração de dados foi usada para calibrar a velocidade dos veículos e, além disso, o trabalho propôs uma nova regressão ponderada localmente com base no agrupamento hierárquico aglomerativo. Primeiro, as amostras de treinamento foram agrupadas e, depois, o algoritmo de agrupamento hierárquico aglomerativo foi usado para estabelecer uma árvore de classificação para cada agrupamento de restrições. Em seguida, o método do *k*-vizinho mais próximo foi empregado para classificar novas amostras no melhor agrupamento ajustado. Finalmente, a velocidade do veículo foi estimada. De acordo com os autores, os testes com uma quantidade enorme de dados de campo demonstraram que o algoritmo proposto teve um bom desempenho na precisão e eficiência da estimativa dos parâmetros.

Kianfar e Edara (2013) pesquisaram a aplicação de técnicas de agrupamento no particionamento dos dados do fluxo de tráfego. Assim, propuseram uma estrutura de agrupamento baseada nas variáveis fundamentais do fluxo de tráfego e compararam três técnicas de agrupamento: Hierárquico, *k-means* e Modelo de Misturas Gaussianas. Os autores aplicaram as técnicas de agrupamento citadas em todas as combinações possíveis formadas pelas variáveis: fluxo, velocidade e ocupação. Segundo eles, os resultados indicaram que o agrupamento é um meio efetivo para classificar os regimes de fluxo dos dados de tráfego visando a criar os diagramas

fundamentais de fluxo de tráfego e os modelos macroscópicos de tráfego. Considerando as combinações entre as variáveis, os melhores resultados do agrupamento foram alcançados usando apenas a velocidade e a ocupação como variáveis de entrada. Além disso, de acordo com o coeficiente *Silhouette* que representa um índice de validação da consistência dos agrupamentos, os pesquisadores destacam que o desempenho das técnicas de agrupamento *k-means* e hierárquico são comparáveis entre si e superam o Modelo de Misturas Gaussianas.

Mais recentemente, Gu *et al.* (2017) propuseram uma metodologia de agrupamento de dois estágios baseados em *big data* para calibrar os diagramas fundamentais de tráfego nas conexões em redes de vias expressas. Basicamente, são empregados o algoritmo *k-means* combinado com o agrupamento hierárquico e o agrupamento hierárquico modificado baseado na distância de *Fréchet* para obter uma matriz de parâmetros fundamentais de tráfego para cada conexão. Depois, o agrupamento hierárquico modificado é reempregado nessas matrizes associado a análise de componentes principais mediante a modificação da medida de similaridade para realçar as variações nos diagramas fundamentais. Finalmente, o agrupamento de séries temporais multivariadas resultante modela as distribuições dos parâmetros dos diagramas fundamentais. A estrutura proposta é aplicada na rede de autoestradas de Melbourne usando dados coletados de um ano. Assim, esses autores concluem que a metodologia proposta ajuda a compreender melhor: a distribuição espacial de conexões com diagramas fundamentais semelhantes, que não representam necessariamente vias com atributos físicos similares, e as variações e distribuições dos parâmetros dos diagramas fundamentais associados.

3 DATA MINING

No curso dos trabalhos citados na revisão bibliográfica deste trabalho, as aplicações do *data mining* sugeriram algumas abordagens específicas mediante técnicas de análises de grandes quantidades de dados ou *big data*, evidenciando alguns conceitos básicos, em especial as técnicas de agrupamento hierárquico aglomerativo e *k-means*, além das métricas de validação da consistência dos *clusters* definidos pelas técnicas de agrupamento. Por isso, tais conceitos são sumariamente apresentados na sequência.

3.1 Agrupamento Hierárquico Aglomerativo

Trata-se de uma abordagem *bottom-up* para formar uma estrutura hierárquica de *clusters*. Basicamente, o algoritmo precisa de uma medida de distância ou de uma medida de similaridade entre dois *clusters*. Assim, cada ponto é inicialmente considerado como um *cluster* independente. Depois, localiza os dois clusters mais próximos que, em seguida, são mesclados. O algoritmo continua nesse processo, registrando todas as fusões de baixo para cima, até restar apenas um *cluster* e formar a sua estrutura hierárquica de agrupamento (Witten *et al.*, 2011, e Tan *et al.*, 2018).

3.2 *k-means*

É a técnica de agrupamento clássica e se baseia na distância iterativa entre os pontos. Primeiro, escolhe-se o parâmetro “*k*” que é o número de classes ou *clusters* em que se pretende agrupar os dados. Depois, *k* pontos são escolhidos aleatoriamente como centro dos clusters e, de acordo com a distância euclidiana, os demais pontos são classificados em relação ao centro do *cluster* mais próximo. Em seguida, o centróide, ou a média, de cada agrupamento é calculado – essa é a parte “*means*” do método. Esses centróides são considerados novos valores centrais para seus respectivos *clusters*. Finalmente, todo o processo é repetido com os novos centros de cluster. A

iteração continua até que os mesmos pontos sejam atribuídos a cada cluster em rodadas consecutivas, estabilizando definitivamente os centroides (Witten *et al.*, 2011, e Tan *et al.*, 2018).

3.3 Coeficiente *Silhouette*

É um método popular que combina a coesão e a separação por meio de três etapas, usando distâncias ou semelhanças, como ensina Tan *et al.* (2018). O coeficiente *Silhouette* médio de um *cluster* é simplesmente definido pela média dos coeficientes *Silhouette* de pontos pertencentes ao *cluster*. A medida geral da qualidade de um agrupamento é obtida pelo coeficiente *Silhouette* médio de todos os pontos. O valor do coeficiente *Silhouette* varia entre -1 e +1, onde o valor +1 indica um agrupamento fortemente consistente e o valor -1 significa que os pontos de dados foram classificados incorretamente.

4 METODOLOGIA

Este trabalho utilizou um banco de dados fornecido pelo departamento de trânsito local – DETRAN – composto pelos registros diários de cada veículo que cruza os laços indutivos instalados nas aproximações de duas interseções urbana. Originalmente, esses registros estão agrupados por meses e cada mês possui em média 520.000 linhas de registros com 07 (sete) informações em cada uma delas, a saber: identificação do equipamento, data, hora, faixa, velocidade registrada, velocidade limite e tamanho do veículo. Conforme citado, são 02 (duas) interseções e cada uma delas possui um sensor para cada uma das 04 (quatro) direções, totalizando 08 (oito) sensores. Os registros contêm 31 (trinta e um) meses de coleta e se estendem de dezembro de 2014 até junho de 2017, totalizando assim 128.960.000 linhas de dados ou 902.720.000 dados para apenas duas interseções.

Portanto, uma enorme quantidade de dados como essa produzida pelos equipamentos de fiscalização eletrônica de veículos tem grande relevância (Stathopoulos e Karlaftis, 2003; Wen, 2008; Karlaftis e Vlahogianni, 2011; García-Nieto *et al.*, 2013) e devem receber o tratamento adequado ao seu vulto (Dunkel *et al.*, 2011; Zeng, 2015; ITF, 2015) para produzir resultados consistentes, oportunos e automáticos diante da dinâmica dos fluxos veiculares (Dervisoglu *et al.*, 2009; Jiang e Huang, 2009; Li e Zhang, 2011; Jiang *et al.*, 2012; Kianfar e Edara, 2013; Gu *et al.*, 2017) no contexto de vias urbanas aproveitando a estrutura de equipamentos de fiscalização nas aproximações de interseções semaforizadas das cidades brasileiras.

Dervisoglu *et al.* (2009) advertem sobre a importância de uma boa medida na performance dos equipamentos de fiscalização eletrônica de veículos no que tange ao tamanho das amostras coletadas e sobre a presença expressiva de dados de fluxo congestionado nessas amostras. Assim, uma análise preliminar dos dados disponíveis para este trabalho constatou que os arquivos foram extraídos do banco de dados do DETRAN mediante a classificação dos registros por data, faixa de trânsito e hora das coletas, enquanto os sensores de fiscalização só ordenam os dados coletados em função do tempo, desconsiderando a ordem das faixas em que a coleta foi registrada. Além disso, foram identificadas defasagens de tempo elevadas entre registros sucessivos que são inconsistentes com os tempos de verdes usuais de um ciclo semaforico. Tais defasagens são caracterizadas por interrupções nos registros com durações entre 10 minutos e 6 horas, aproximadamente, sugerindo a ocorrência de quedas de energia, defeitos e atividades de manutenção. Sendo, portanto, necessário realizar um tratamento prévio nos dados coletados conforme é apresentado na sequência. A Figura 1 exemplifica algumas lacunas em uma das séries temporais registradas pelos sensores.

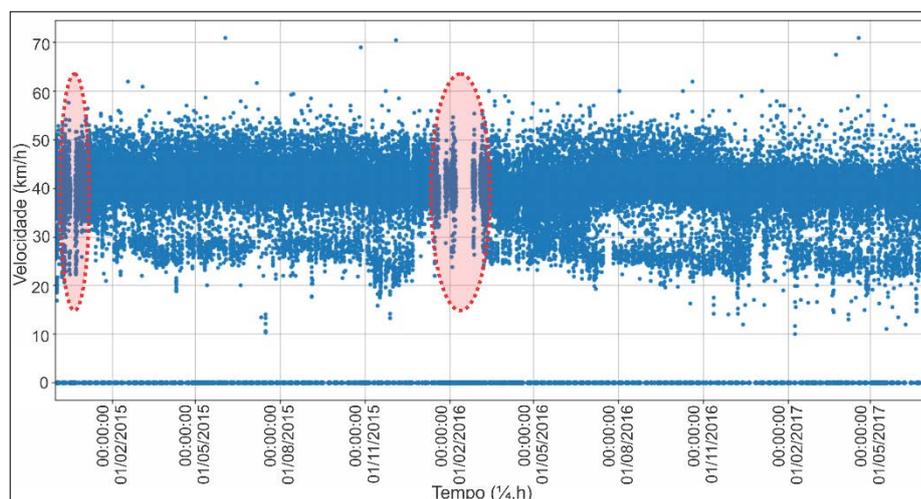


Figura 1: Lacunas na série temporal de velocidades registradas por um dos sensores.

4.1 Tratamento dos Dados Coletados

Preliminarmente, um algoritmo foi desenvolvido para reunir os arquivos em um arquivo único, reordenar esses arquivos de acordo com o momento de cada coleta e desconsiderar as interrupções a fim de minerar os dados. Nesse processo, o algoritmo acumulou os fluxos e as velocidades em intervalos de tempo de 15 minutos. A partir desses dados, a mineração produziu as séries temporais de fluxos médios e de velocidades médias para esses intervalos. Observando que os fluxos das três faixas de rolamento foram agrupados em um único fluxo nesta análise preliminar. Depois, esses gráficos foram discretizados em períodos semanais. A Figura 2 exemplifica uma das séries temporais que resultou dessa etapa e desenvolve os fluxos e velocidades a partir de uma segunda-feira (12/09/2016) até um domingo (18/09/2016), pois, para fins de avaliação do comportamento do tráfego no final de semana, ficou mais adequado juntar o sábado e o domingo. Finalmente, as informações semanais foram refinadas em dias de semana.

A Figura 3 exemplifica uma das séries temporais que resultou dessa última etapa e desenvolve os fluxos e velocidades para uma quarta-feira (14/09/2016) e uma quinta-feira (15/09/2016), indicando o comportamento dessas variáveis nos horários de pico matutino por meio das linhas vermelhas. Assim, nesta análise preliminar dos dados, ficaram evidenciados por meio da interpretação dos gráficos produzidos os períodos desfavoráveis, bem como o melhor horário ao longo do dia para aplicação da modelagem. Assim, foram escolhidos os meses de maio, junho, julho e agosto do ano de 2016 e o período compreendido entre as 05:00 e as 10:00h de terça a quinta-feira de cada semana para o estudo dos gráficos fundamentais de tráfego e a calibração dos respectivos parâmetros (fluxo, velocidade e densidade).

4.2 Modelagem e Calibração

A bibliografia citada na revisão da literatura deste artigo demonstra que o *data mining*, apoiado nas técnicas de agrupamento, tem boa aderência na classificação dos fluxos de trânsito, principalmente devido ao emprego das técnicas *k-means* e agrupamento hierárquico. A consistência da técnica de agrupamento *k-means*, inclusive as de suas adaptações com outros métodos, induz à consolidação do seu emprego como ferramenta de classificação desse fluxo de acordo com os seus estados possíveis. Assim, foi desenvolvido outro algoritmo na linguagem *Python* para analisar os dados previamente tratados.

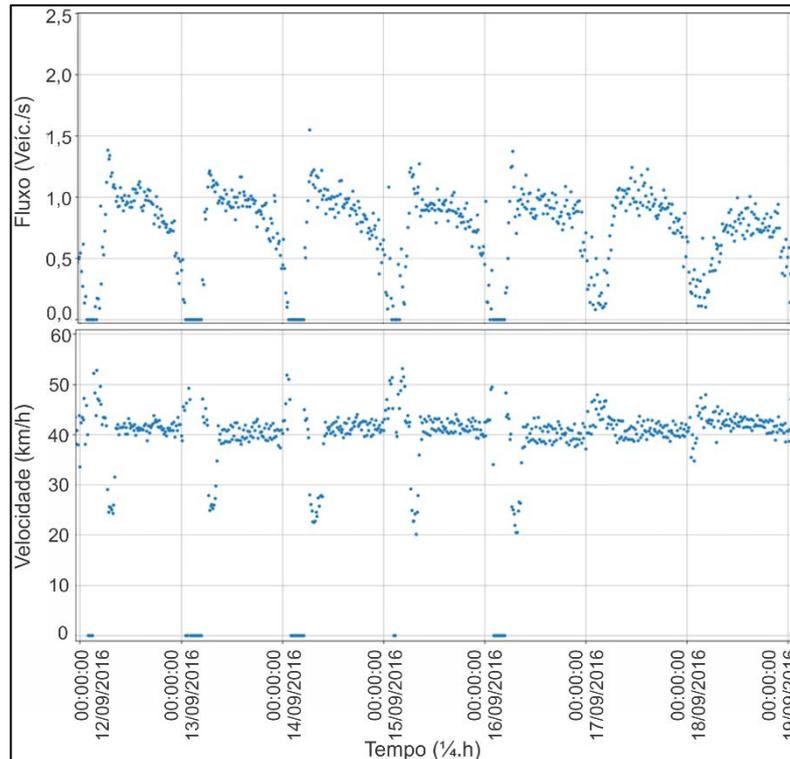


Figura 2: Série temporal diária de fluxo e velocidade médias para intervalos de 15 minutos.

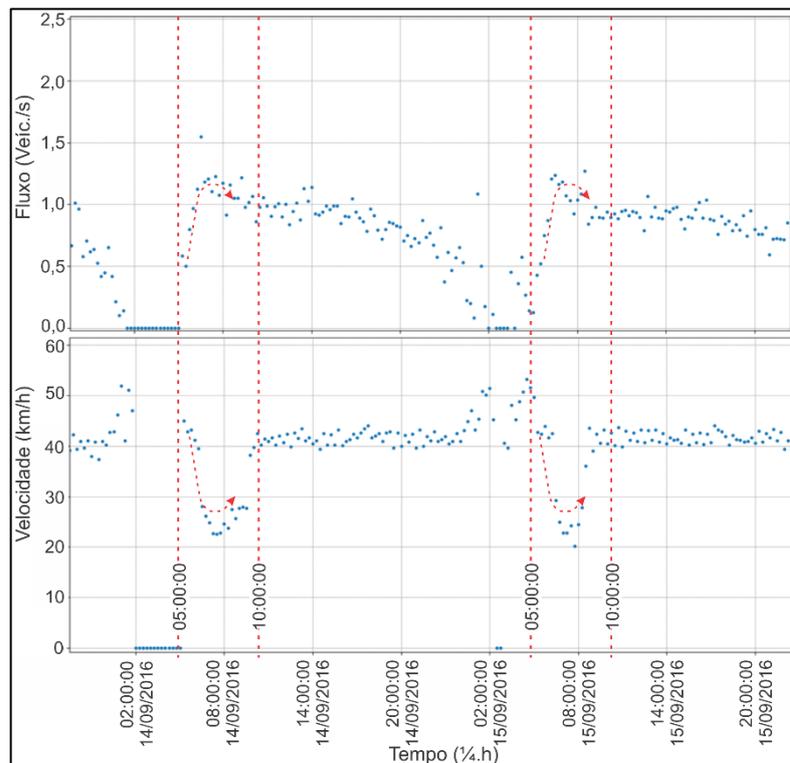


Figura 3: Série temporal horária de fluxo e velocidade médias para intervalos de 15 minutos.

De forma sucinta, esse algoritmo acumula os dados de fluxo em função do tempo, calcula a velocidade média e determina a densidade do tráfego no mesmo período e forma os pontos sob a forma de pares ordenados (Figura 4a). Depois, faz a classificação dos pontos de acordo com

o método *k-means* e as variáveis escolhidas a fim de formar os *clusters* desses pontos conforme o estado em que foram classificados. A Figura 4 compara o mesmo conjunto de dados de acordo com dois cenários distintos: sem qualquer classificação (a) e com *clusters* ou agrupamentos (b). Esses *clusters* representam cada um dos dois estados possíveis do fluxo em apreço, ou seja, eles agrupam os dados com características de fluxo livre e de fluxo congestionado, criando um limite entre esses dois estados. Os agrupamentos foram realizados por meio da variável densidade e, também, pela sua combinação com a variável velocidade, a fim de gerar os gráficos fundamentais: Densidade x Fluxo, Velocidade x Fluxo e Velocidade x Densidade com pontos classificados.

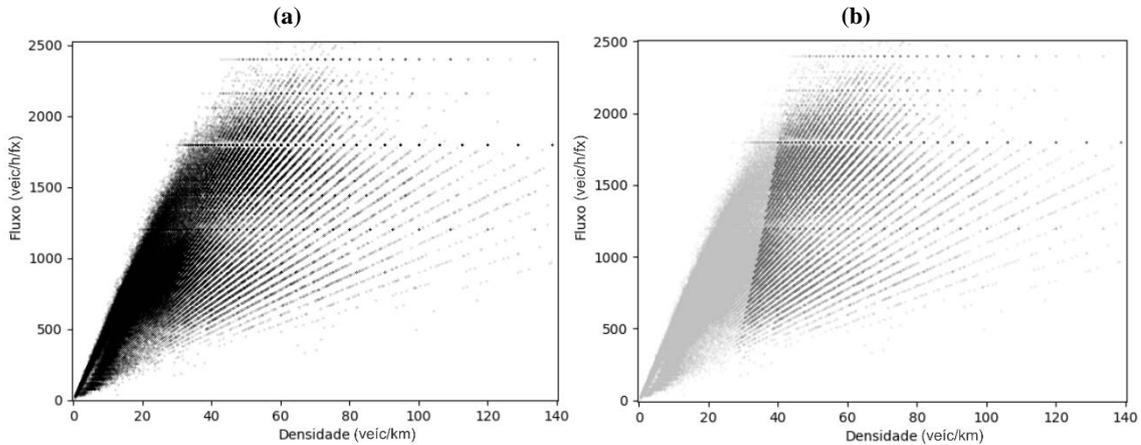


Figura 4: Comparação entre os gráficos Densidade x Fluxo sem (a) e com agrupamentos (b).

Até onde se tem conhecimento as pesquisas sobre a calibração de parâmetros de tráfego de veículos foram realizadas com base em dados coletados em seções livres de rodovias, cujo comportamento se caracteriza predominantemente por faixas elevadas de fluxos e velocidades. Entretanto, quando a coleta é realizada nas aproximações de uma interseção semaforizada, uma determinada faixa de valores de fluxo ou velocidade não se destaca, pois os veículos transitam em diversas configurações de tráfego, inclusive partindo da inércia. Assim, por exemplo, não se pode esperar um cone estreito para representar o fluxo livre no gráfico fundamental Densidade x Fluxo como ocorre nos trabalhos de Dervisoglu *et al.* (2009) e Kianfar e Edara (2013), mas sim um arranjo diferente e mais amplo dessas variáveis, desde os valores bem menores até os mais elevados, formando uma nuvem de pontos mais espalhada em todo o espectro de velocidades permitidas na via. Por isso, em seguida, uma regressão linear foi ajustada a cada um dos fluxos de tráfego, definindo uma interseção entre as duas retas resultantes (Figura 5). Essa interseção define a velocidade em que ocorre a transição de fluxo do estado livre para o estado congestionado.

Ora, tomando-se o gráfico Densidade x Fluxo, têm-se que qualquer reta que passe pela origem de seus eixos representa uma dada velocidade que é determinada pelo coeficiente angular da respectiva reta (Equação 1), ou seja, tomando a densidade (ρ), o fluxo (q) e a velocidade (u):

$$se, \rho = q/u \rightarrow q(\rho) = u \cdot \rho \quad (1)$$

Assim, a reta, que passa pela origem do sistema de eixos Densidade x Fluxo e pela interseção das regressões lineares ajustadas aos dados de fluxos livre e congestionado, representa a velocidade de transição entre esses estados de fluxo, ou seja, a velocidade crítica (Figura 6). Como a fronteira entre os *clusters* define a mudança de características entre os fluxos livre e congestionado, o ponto de interseção entre essa fronteira e a reta da velocidade crítica define o ponto crítico do fluxo e da densidade também.

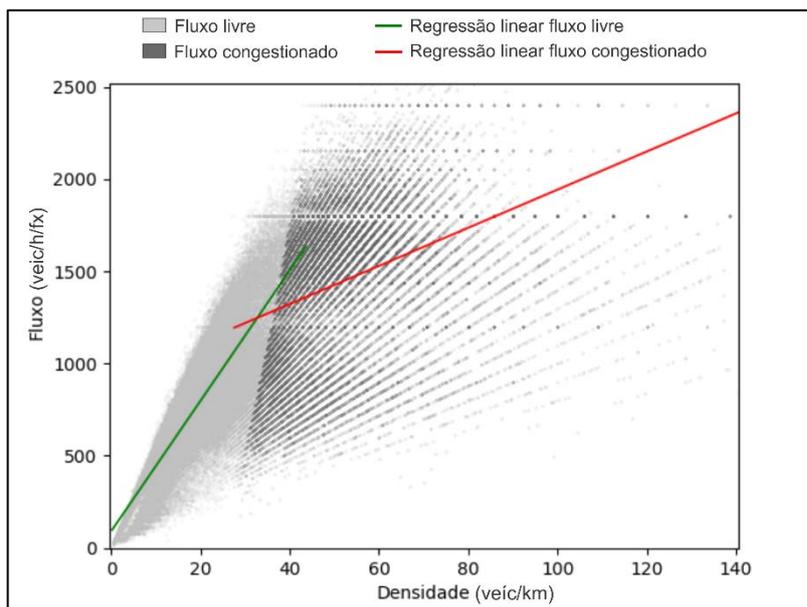


Figura 5: Regressões lineares ajustadas aos fluxos livre e congestionado

Nesse caso, é possível extrair os valores críticos das variáveis pela projeção do ponto crítico nos respectivos eixos. A Figura 7 representa essa solução analítica, definindo o ponto crítico e suas projeções sobre o sistema de eixo.

5 RESULTADOS OBTIDOS

Os resultados relevantes são obtidos a partir do agrupamento e, conseqüentemente, se devem à medida de performance da aplicação do método *k-means*, ou seja, ao quanto a classificação adere às características de cada ponto. Neste trabalho, essa validação foi realizada por meio do coeficiente *Silhouette* aplicado a cada *cluster* de acordo com todas as 7 combinações possíveis entre as variáveis densidade, fluxo e velocidade. A Tabela 1 apresenta esses resultados. É possível observar que todos os valores são próximos entre si. Entretanto, existe uma diferença considerável entre os coeficientes *Silhouette* relativos às variáveis densidade e velocidade. Assim, observa-se também, considerando os *clusters* formados mediante as combinações possíveis com a variável densidade: densidade (0,677), densidade com fluxo (0,664) e densidade com velocidade (0,654), que a densidade conduziu classificações mais consistentes do que aquelas realizadas a partir do fluxo ou da velocidade.

Neste modelo proposto, somente a densidade produz agrupamentos com fronteiras com comportamentos verticais, enquanto o fluxo e a velocidade produzem respectivamente fronteiras com comportamentos horizontais e inclinados. Com fronteiras horizontais, a convergência do modelo se dá muito além dos parâmetros de tráfego aceitáveis e, com fronteiras inclinadas, o modelo não converge. Portanto, considerando que o modelo se baseia na convergência de comportamentos críticos do tráfego, busca-se uma fronteira vertical ou que se aproxime da vertical. De todas as combinações possíveis entre as variáveis de classificação dos agrupamentos de tráfego livre ou congestionado, apenas duas combinações produzem fronteiras com comportamentos verticais: densidade e densidade com velocidade. A Figura 8 compara os resultados das classificações para cada uma das 3 variáveis: densidade, fluxo e velocidade, a fim de caracterizar o comportamento das fronteiras resultantes.

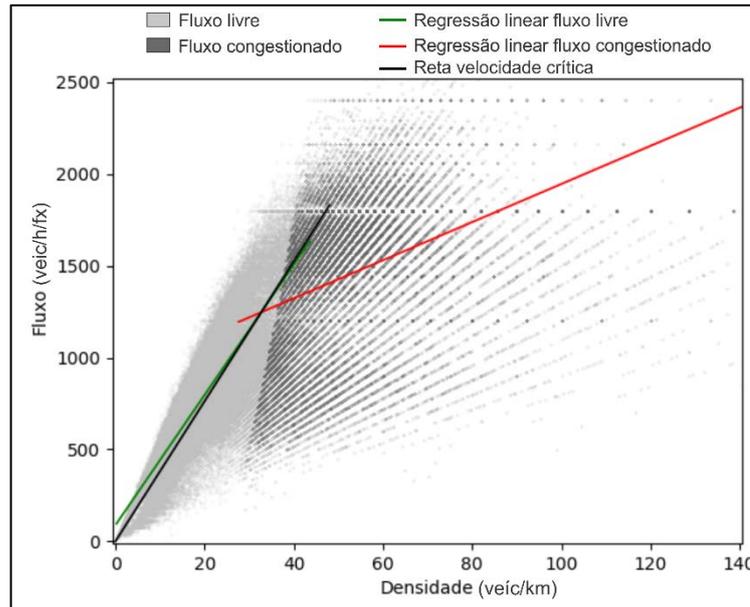


Figura 6: Determinação da velocidade crítica

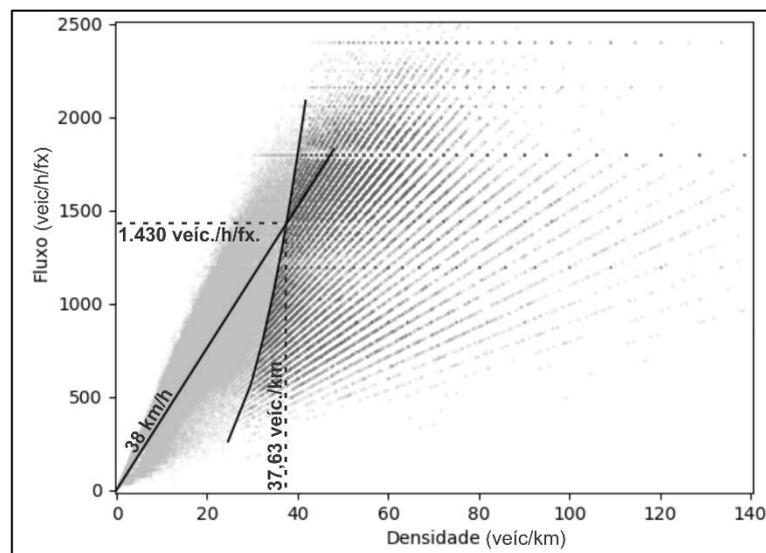


Figura 7: Identificação do ponto crítico e determinação de suas projeções no sistema de eixos.

Tabela 1: Coeficientes *Silhouette* para cada combinação entre as variáveis de classificação.

VARIÁVEIS	Densidade	Fluxo	Velocidade
Densidade	0,677	0,664	0,654
Fluxo	-	0,662	0,587
Velocidade	-	-	0,566
Todas		0,660	

Então, considerando o melhor desempenho da densidade na classificação dos fluxos e a característica das fronteiras produzidas por meio desta variável, a Tabela 2 destaca os resultados obtidos pelo modelo para as duas combinações possíveis. Além disso, a Tabela 2 apresenta o percentual do fluxo calculado em relação a uma capacidade de 1.800 veículos por hora por faixa.

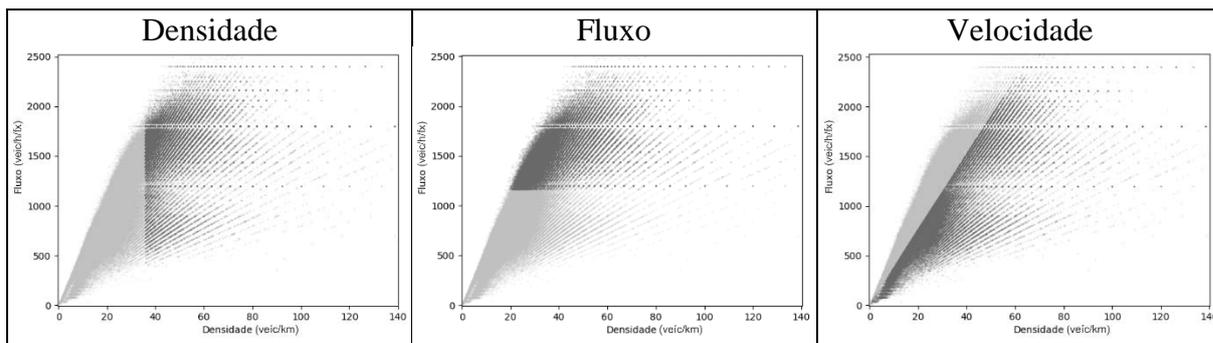


Figura 8: Comparação entre os comportamentos das fronteiras dos *clusters* de acordo com os parâmetros de classificação.

Tabela 2: Parâmetros fundamentais calibrados automaticamente em função das variáveis de classificação.

Variáveis de Clusterização	Fluxo (veic./h/fx.)	Densidade (veic./km)	Velocidade (km/h)	% Capacidade	Coefficiente <i>Silhouette</i>
Densidade	1.374,38	36,20	37,97	76,35	0,677
Densidade e Velocidade	1.430,00	37,63	38,00	79,45	0,587

6 DISCUSSÕES E CONCLUSÕES

Embora os resultados dos agrupamentos apresentados na Figura 8 possam sugerir que a simples sobreposição das fronteiras de transição entre os fluxos livre e congestionado de cada gráfico possa indicar uma solução adequada para calibrar os parâmetros fundamentais de tráfego, cabe destacar que o objetivo é propor uma metodologia aplicável em tempo real a uma quantidade enorme de dados. E, por isso, quanto menor a quantidade de operações melhor o desempenho da metodologia. A sobreposição citada requer a formação e a classificação dos pontos para a construção dos três gráficos fundamentais enquanto a metodologia proposta neste trabalho requer apenas as operações necessárias para produzir um gráfico.

Para o caso estudado neste trabalho, os resultados indicados pela Tabela 2 são bastante coerentes. Cabe destacar que diversas publicações técnicas indicam com base no HCM (2010) que 1.800 veículos por hora por faixa é a capacidade máxima de vias arteriais e coletoras urbanas no limite a montante da interseção segundo condições ideais de geometria e pavimentação. Ora, se de acordo com Pietrantonio (2017), dados empíricos indicam que o fluxo congestionado ocorre entre 70 e 80% da capacidade da via, então, os fluxos calculados pela metodologia de calibração proposta se enquadram nesses limites e caracterização a transição entre o fluxo livre e o fluxo congestionado.

Finalmente, ao contrário de trabalhos anteriores, este trabalho utiliza espectros de velocidades e fluxos muito amplos e, portanto, variados, pois se dedica à análise do tráfego urbano nas aproximações de uma interseção semaforizada em que os veículos podem partir da inércia ou transitar em marcha contínua livre ou forçada. Então, os coeficientes *Silhouette* calculados demonstram um agrupamento muito consistente e também a aptidão da variável densidade como variável de classificação, o que se alinha com os resultados de trabalhos anteriores. Desta forma, ainda que não existam trabalhos iguais a este, a metodologia proposta aqui para a calibração automática de parâmetros fundamentais de tráfego se demonstra bastante promissora, bem como uma excelente opção em relação ao palpite dos pesquisadores e às campanhas para levantamentos de dados em campo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Chen, S. e D. J. Sun (2016) An Improved Adaptive Signal Control Method for Isolated Signalized Intersection Based on Dynamic Programming. *IEEE Intelligent Transportation Systems Magazine*, v. 8, n. 4, p. 4-14.
- Dervisoglu, G., G. Gomes, J. Kwon; R. Horowitz e P. Varaiya (2009) Automatic Calibration of The Fundamental Diagram and Empirical Observations on Capacity. *Transportation Research Board 88th Annual Meeting*, TRB, Washington/DC, USA, v. 15.
- Dunkel, J.; A. Fernández; R. Ortiz e S. Ossowski (2011) Event-Driven Architecture for Decision Support in Traffic Management Systems. *Expert Systems with Applications*, v. 38, n. 6, p. 6530-6539.
- García-Nieto, J.; A. C. OLIVERA e E. Alba (2013) Optimal Cycle Program of Traffic Lights with Particle Swarm Optimization. *IEEE Transactions on Evolutionary Computation*, v. 17, n. 6, p. 823-839.
- Gu, Z.; Saberi, M.; Sarvi, M. e Liu, Z. (2017) A Big Data Approach for Clustering and Calibration of Link Fundamental Diagrams for Large-Scale Network Simulation Applications. *Transportation Research Part C: Emerging Technologies*.
- Guan, D. e Z. Yang (2002) Signal Timing Optimization for Urban Traffic Adaptive Control System. *International Conference on Traffic and Transportation Studies 2002*, ASCE, Guilin, China, p. 871-876.
- Hand, D. J.; H. Mannila e P. Smyth (2001) *Principles of Data Mining*. Ed. MIT Press, Cambridge, MA, USA.
- HCM (2000) Highway Capacity Manual. Transportation Research Board, Washington, DC, USA.
- HCM (2010) Highway Capacity Manual. Transportation Research Board, Washington, DC, USA
- ITF (2015) Big Data and Transport: Understanding and assessing options. OECD Corporate Partnership Board, Paris, France.
- Jiang, Z. e Y. X. Huang (2009) Parametric Calibration of Speed-Density Relationships in Mesoscopic Traffic Simulator with Data Mining. *Information Sciences*, v. 179, n. 12, p. 2002-2013.
- Jiang, Z.; L. I. Shubin e L. I. U. Xiaoqing (2012) Parameters Calibration of Traffic Simulation Model Based on Data Mining. *Journal of Transportation Systems Engineering and Information Technology*, v. 12, n. 6, p. 28-33.
- Karlaftis, M. G. e E. I. Vlahogianni (2011) Statistical Methods Versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights. *Transportation Research Part C: Emerging Technologies*, v. 19, n. 3, p. 387-399.
- Kianfar, J. e P. Edara (2013) A Data Mining Approach to Creating Fundamental Traffic Flow Diagram. *Procedia-Social and Behavioral Sciences*, v. 104, p. 430-439.
- Li, J. e H. M. Zhang (2011) Fundamental Diagram of Traffic Flow: New Identification Scheme and Further Evidence from Empirical Data. *Transportation Research Record*, v. 2260, n. 1, p. 50-59.
- Luna, M. S. (2003) Sobre o Fluxo de Saturação: Conceituação, Aplicação, Determinação e Variação. *Dissertação de Mestrado*, Programa de Mestrado em Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, CE.
- Pietrantonio, H. (2017) Engenharia de Tráfego: Introdução à Teoria do Fluxo de Tráfego. *Escola Politécnica da Universidade de São Paulo – Departamento de Engenharia de Transportes*. Disponível em: <<http://sites.poli.usp.br/d/ptr5803/ET2-Teoria.pdf>>. Acesso em: 01 fev. 2017
- Stathopoulos, A e M. G. Karlaftis (2003) A multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction. *Transportation Research Part C: Emerging Technologies*, v. 11, n. 2, p. 121-135.
- Sun, L. e J. Zhou (2005) Development of Multiregime Speed-Density Relationships by Cluster Analysis. *Transportation Research Record*, v. 1934, p. 64-71.
- Tan, M.; S. C. Wong; J. Xu; Z. Guan e P. Zhang *et al.* (2007) An Aggregation Approach to Short-Term Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, v. 10, n. 1, p. 60-69.
- Tan, P.; Steinbach, M., Karpatne, A. e Kumar, V. (2018) *Introduction to data mining*. (2ª ed.). Ed. Pearson, London, England.
- Wen, W. (2008) A Dynamic and Automatic Traffic Light Control Expert System for Solving The Road Congestion Problem. *Expert Systems with Applications*, v. 34, n. 4, p. 2370-2381
- Witten, I. H.; E. Frank; M. A. Hall e C. J. Pal (2011) *Data Mining: Practical machine learning tools and techniques* (3ª ed.). Ed. Morgan Kaufmann, Burlington, MA, USA.
- Zeng, G. (2015) Application of Big Data in Intelligent Traffic System. *IOSR Journal of Computer Engineering*, v. 17, n. 1, p. 01-04.

Fábio de Souza Pereira Borges (fabiospb@gmail.com)

Adelaida Pallavicini Fonseca (ixcanil@unb.br)

Li Weigang (weigang@unb.br)