

CERVA: Roteamento Contextual para Veículos com Risco Espaço-temporal

Lucas Zanco Ladeira^{1,2}, Allan Mariano de Souza¹, Thiago Henrique Silva³,
Geraldo Pereira Rocha Filho⁴, Maycon Leone Maciel Peixoto^{1,5}, Leandro Aparecido Villas¹

¹Laboratório de Redes de Computadores, Universidade Estadual de Campinas,
Cidade Universitária, Campinas, SP, Brasil

²CPQD, Centro de Pesquisa e Desenvolvimento em Telecomunicações,
Parque II do Polo de Alta Tecnologia, Campinas, SP, Brasil

³Departamento de Informática, Universidade Tecnológica Federal do Paraná
Av. 7 de Setembro, 3165, Curitiba, PR, Brasil

⁴Departamento de Ciência da Computação, Universidade de Brasília,
Campus Darcy Ribeiro, Brasília, DF, Brasil

⁵Departamento de Ciência da Computação,
Universidade Federal da Bahia (UFBA), Salvador – Bahia – Brasil
{lucaszl, allanms}@lrc.ic.unicamp.br, thiagoh@utfpr.edu.br,
geraldof@unb.br, maycon.leone@ufba.br, lvillas@unicamp.br

Abstract. *Nowadays, there is a shortage of real mobility data openly available. Thus, several works in the literature generate synthetic mobility, which does not represent real mobility. Some of these works use contextual data to propose route recommendations but do not study the behavior of such data. In addition, the impact of each contextual data type changes according to the user's profile. To solve the problems mentioned above, CERVA is proposed, a contextual routing solution for vehicles with space-time risk. CERVA consists of three modules: contextual window identification, context mapping, and routing customization. The evaluation results show that CERVA recommends the best routes according to the user's profile.*

Resumo. *Nos dias atuais existe uma escassez de dados de mobilidade reais disponíveis abertamente. Sendo assim, diversos trabalhos da literatura geram mobilidade sintética a qual não representa a mobilidade real. Alguns desses trabalhos fazem o uso de dados contextuais para propor recomendação de rotas, no entanto não estudam o comportamento de tais dados. Além disso, o impacto de cada tipo de dado contextual muda de acordo com o perfil do usuário. Para resolver os problemas citados anteriormente é proposto o CERVA, uma solução de roteamento contextual para veículos com risco espaço-temporal. O CERVA é composto por três módulos, sendo: identificação das janelas contextuais, mapeamento de contexto, e personalização do roteamento. Os resultados da avaliação mostram que o CERVA recomenda as melhores rotas de acordo com o perfil do usuário.*

1. Introdução

O crescimento demográfico nas cidades pode implicar em um maior fluxo de veículos. Isso vem contribuindo para tornar congestionamentos mais frequentes,

reduzindo a segurança nas estradas e tornando o gerenciamento de recursos problemático [Sahana et al. 2018]. Nesse contexto, várias aplicações tentam melhorar a mobilidade urbana, tais como sincronização de semáforos, monitoramento de ônibus em tempo real e identificação de estradas congestionadas. Essas aplicações representam uma categoria de sistemas denominada *Intelligent Transport Systems* (ITS). De acordo com Cunha et al. [Cunha et al. 2017], o ITS busca resolver problemas relacionados ao transporte nas cidades e, para isso, utiliza dados referentes ao tráfego de veículos, bem como novas tecnologias para obter e disseminar dados.

Existem diversos trabalhos da literatura na área de ITS que utilizam a mobilidade veicular dentro de simuladores para validar protocolos de comunicação em VANETs [Liu et al. 2018, Akabane et al. 2018, Ghorai and Banerjee 2017], alocação de recursos [Santos et al. 2016, Jaimes et al. 2016], e algoritmos de sugestão de rotas [Galbrun et al. 2016, Akabane et al. 2018, Souza et al. 2018]. Os trabalhos propostos geralmente filtram bases reais que são disponibilizadas^{1,2,3}. Entretanto, tais bases são escassas não possibilitando aplicar em cenário específicos. Sendo assim, uma alternativa é gerar rotas aleatoriamente por ferramentas como RandomTrips⁴. Salienta-se, entretanto, que essas rotas aleatórias não representam a mobilidade real. Alguns serviços *Freemium*⁵ (e.g., Here API) disponibilizam dados reais sobre o fluxo de tráfego nas vias. Com esses dados é possível analisar o comportamento do tráfego de acordo com horas distintas no dia, e gerar padrões de mobilidade sintéticos mais próximos da realidade. Isso remove a restrição da quantidade de bases reais disponíveis, e permite explorar um maior conjunto de cenários nas simulações.

Nesse cenário, dados contextuais podem ser utilizados para agregar na tarefa de sugerir rotas, auxiliando na obtenção novos tipos de serviços. Os dados contextuais relacionados à dinâmica da cidade podem ser utilizados, por exemplo, para identificar os níveis de segurança de uma via. Entendemos que o comportamento, isto é, a distribuição ao longo do dia, de dados contextuais específicos pode ter padrões úteis se capturados e analisado adequadamente. Isso ocorre porque o comportamento dos dados contextuais muda regularmente de acordo com o período do dia [Newton and Felson 2015, Malleson and Andresen 2015, Andresen and Malleson 2015, Felson and Boivin 2015]. Por exemplo, uma região não segura determinada pode não apresentar o mesmo comportamento em um período diferente do dia. Se esse for o caso, é possível identificar o intervalo de tempo em que os dados contextuais se comportam de maneira semelhante, as chamadas janelas contextuais deste estudo.

Janelas contextuais são úteis para vários serviços que o ITS busca resolver. Por exemplo, se identificarmos janelas relacionadas à segurança, ajudará a fornecer rotas seguras em áreas específicas da cidade ou em determinados períodos do dia. Cada tipo de dados contextuais pode ter um impacto distinto de acordo com o perfil de usuários. No exemplo de segurança, alguns usuários podem querer evitar todos os tipos de riscos, enquanto outros podem evitar apenas aqueles que são mais perigosos e não afetam significativamente o tempo de viagem. Consequentemente, personalizar um serviço de redirecionamento é um recurso essencial e não trivial de resolver para melhorar a experiência do usuário.

¹<https://data.austintexas.gov/>

²<https://data.cityofchicago.org/>

³<https://sumo.dlr.de/docs/Data/Scenarios/TAPASCologne.html>

⁴<https://sumo.dlr.de/docs/Tools/Trip.html>

⁵Tipo de serviço que disponibiliza gratuitamente o uso até uma quantidade de requisições e/ou subconjunto de ferramentas, e então, cobra um valor pelo excedente.

Com o objetivo de resolver os problemas mencionados anteriormente, como de obter um maior conjunto de dados de mobilidade, identificar o comportamento dos dados e permitir personalização do roteamento, este trabalho propõe uma estratégia que utiliza dados reais de fluxo de tráfego para gerar padrões de mobilidade e um serviço de sugestão de rotas personalizadas chamado CERVA. O CERVA utiliza tipos de dados contextuais distintos para identificar janelas de contexto que devem ser evitadas de acordo com a preferência dos usuários. Para tal, é proposta uma estratégia para processar esses tipos de dados baseada em trabalhos anteriores [Ladeira et al. 2019a], auxiliando a determinação de janelas contextuais que compreendem cada situação em períodos distintos do dia. Finalmente, a personalização das rotas compreende um peso atribuído a cada tipo de dado contextual e é usado durante a tarefa de sugestão. Dados históricos disponíveis abertamente e simulações da mobilidade são usados para validar a proposta. Os resultados das simulações mostram que o CERVA recomenda as melhores rotas de acordo com o perfil de um usuário.

O restante deste trabalho está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 apresenta como o CERVA foi desenvolvido. A Seção 4 descreve a metodologia de avaliação e os resultados obtidos. Por fim, a Seção 5 apresenta as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

Há vários trabalhos na literatura que apresentam métodos para recomendação de rotas por meio da utilização de índices de fluxo de tráfego [Zheng et al. 2016, Souza et al. 2019, Rezaei et al. 2017, Akabane et al. 2018]. Esses trabalhos procuram melhorar o tempo de percurso, recomendando as rotas mais rápidas. Esses trabalhos consideram apenas informações relacionadas ao tráfego, por exemplo, o fluxo de veículos nas vias. Entretanto, as rotas rápidas sugeridas podem permitir que os veículos passem por regiões perigosas, colocando os usuários em risco. Algumas soluções, tais como Google Maps e Waze, são capazes de sugerir rotas mais rápidas por meio da detecção de níveis de congestionamento de tráfego em tempo-real, como também utilizam dados de segurança fixos para evitar locais inseguros. Assim, surge a necessidade por soluções que realizem a recomendação de rotas seguras por meio da utilização de informações contextuais que incluam a segurança. Essas informações sobre a segurança de uma região podem ser compostas por relatórios sobre as atividades criminais, dados de redes sociais e sistemas de sensoriamento participativo.

Os trabalhos [Shah et al. 2011, Elsmore et al. 2014, Secron et al. 2016] propõem sistemas que são baseados em *crowdsourcing* para utilização de informações sobre atividades criminais reportadas pelos usuários para sugerir rotas mais seguras. Já [Kim et al. 2014], propõe um sistema que recomenda rotas seguras baseadas nas informações e percepções de sentimentos dos usuários do Twitter sobre uma localização. Em [Utamima and Djunaidy 2017], é proposta uma aplicação baseada em Web que utiliza uma API do Google para informar e propor novas rotas seguras aos usuários de uma determinada região. Entretanto, os trabalhos citados anteriormente não consideram o fluxo de veículos nas vias para as rotas seguras recomendadas, podendo piorar a eficiência do tráfego em áreas consideradas com maior nível de segurança. Além disso, tais trabalhos possuem uma base de conhecimento sobre crimes possivelmente incompleta e desatualizada em função das informações que são limitadas aos usuários do próprio sistema.

Já [Keler and Mazimpaka 2016], propõe a caracterização de área inseguras por meio da utilização de informações de *crowdsourcing*, mas combinado a dados abertos do departamento de polícia sobre históricos de crime da região. Os autores mapeiam as ativi-

dades criminais como obstáculos que devem ser evitados no momento da recomendação de rotas seguras. Entretanto, além de apresentar o mesmo problema dos trabalhos citados anteriormente, que não consideram o tráfego dos veículos, esse trabalho não é aplicável em muitos cenários em que não é possível evitar completamente as áreas inseguras por considerá-las como um obstáculos.

Em [Galbrun et al. 2016], os autores utilizam dados criminais públicos com origem em uma iniciativa do governo dos Estados Unidos⁶, combinados a um modelo de mobilidade que inclui a distância percorrida por pedestres das cidades de Chicago e Filadélfia. Eles utilizam a soma de todas as probabilidades de uma *Gaussiana Kernel Density Estimation* (KDE) para caracterizar a densidade criminal de uma região e obter a rota mais segura. Além disso, eles consideram a melhor solução como uma função multi-objetivo que ponderam tanto distância como a segurança. A primeira limitação desse trabalho está relacionada ao custo computacional, por necessitar processar uma grande quantidade de rotas. Outra limitação é a utilização da mobilidade de pedestre no lugar da real mobilidade veicular, podendo direcionar os veículos para vias congestionadas na intenção de evitar áreas com riscos.

Ambos os trabalhos em [Santos et al. 2018, Souza et al. 2019] sugerem rotas alternativas por meio da análise contextual dos níveis de risco de segurança das vias e condições do tráfego. Nesses trabalhos, são aplicadas técnicas de clusterização sobre os dados de crimes. Esses dados sobre atividades criminais são oriundos de fontes abertas e servem para a recomendação da rota segura. Enquanto o trabalho de [Santos et al. 2018] considera o tempo de vida dos dados, o trabalho de [Souza et al. 2019] não inclui a representação da janela temporal de incidência dos dados criminais, ou seja a relação de migração da criminalidade ao longo do tempo. Entretanto, ambos os trabalhos representam as áreas afetadas pelas atividades criminais como regiões circulares e não vias específicas.

Em trabalhos prévios [Ladeira et al. 2019a, Ladeira et al. 2019b] são propostos métodos de representação das atividades criminais nas vias. O trabalho [Ladeira et al. 2019a] se especializa em identificar as flutuações espaço-temporais dos dados criminais, mas não é aplicado diretamente na sugestão de rotas. Por outro lado, o trabalho em [Ladeira et al. 2019b] realiza a identificação de clusters de criminalidade em janelas fixas de 1 hora. Essa informação é utilizada para encontrar rotas mais rápidas, mais curtas, e mais seguras em um *pool* de rotas disponíveis.

Este trabalho se difere das pesquisas apresentadas anteriormente por apresentar uma nova abordagem de mapeamento de mobilidade que caracteriza o fluxo de tráfego por meio de dados reais fornecidos pelo Here. Além disso, foi proposto um novo mecanismo na ponderação dos índices que compõem o *rank sum*[Gunantara 2018] do algoritmo de recomendação de rotas, chamado *Weighted-sum*. Essa evolução do algoritmo permite maior controle e estabilidade na variação dos valores de pesos.

3. CERVA: roteamento Contextual para Veículos com risco espaço-temporal

Esta seção apresenta o CERVA, um serviço que utiliza dados reais contextuais e de fluxo de tráfego para recomendar rotas seguras. A Figura 1 ilustra o funcionamento de cada módulo do CERVA. No Rótulo A da Figura 1, é realizada a identificação das janelas por meio do comportamento dos dados contextuais. Já no Rótulo B da Figura 1 é gerado o mapeamento do contexto com base em KDEs distintas (*Kernel Density Estimation*). A

⁶<http://www.data.gov/about>

ideia é relacionar com maior precisão o risco de cada tipo de dado com as vias as quais os veículos serão re-roteados. Por fim, no Rótulo C da Figura 1 é realizada a personalização do roteamento que permite aos usuários configurarem quanto cada tipo de dado contextual contribui para o risco, de acordo com a sua percepção.

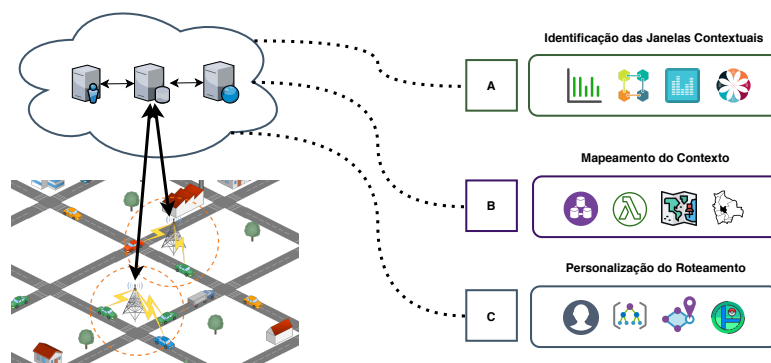


Figura 1. CERVA - Cenário de funcionamento do serviço proposto.

O CERVA possui uma arquitetura centralizada de forma que possa ser utilizado em um servidor em nuvem, recebendo requisições dos veículos que desejam obter rotas alternativas. No CERVA, assume-se que as requisições são transmitidas por meio da infraestrutura LTE. O CERVA identifica o risco de cada via através de dados abertos da cidade de Chicago⁷, e o índice de tráfego pela API do Here⁸. O risco se refere aos dados de ocorrências criminais e acidentes envolvendo veículos, os quais apresentam variáveis espaciais e temporais. Para obter o índice de tráfego foi necessário mapear os dados obtidos pela API no cenário da simulação. Com o intuito de deixar claro ambas as contribuições do trabalho, primeiro será apresentada a estratégia para mapear os dados reais de fluxo de tráfego na Seção 3.1. Em seguida, na Seção 3.2 será apresentado o módulo da identificação das janelas contextuais. Subsequentemente, será descrito o módulo de mapeamento do contexto na Seção 3.3. Por fim, na Seção 3.4, será apresentado o módulo de personalização de rotas.

3.1. Mapeamento do Tráfego

Para mapear os dados reais de tráfego, foi utilizada a Traffic API do serviço Here⁹ que além de obter os dados de tráfego dos veículos, permite requisitar dados de tráfego considerando uma região geográfica definida. Essa API dispõe de um zoom semântico para uma melhor precisão dos dados de acordo com o aumento do zoom. O retorno da API é no formato XML contendo o índice de tráfego, o nome da rua, o código da rua no formato *Point TMC Location Code*, a velocidade média no momento, entre outros dados. Com esses dados é possível verificar diferentes aspectos do cenário do fluxo do tráfego no momento da requisição.

Os dados utilizados para simulação são obtidos pela ferramenta *OpenStreetMap* (formato *.osm*), e então são convertidos para o formato *.net.xml*, formato compatível com o simulador Simulator of Urban Mobility (SUMO). Em ambos os casos o cenário é tratado como um grafo, considerando as vias como arestas, e os nós para conectar vias. A

⁷<https://data.cityofchicago.org>

⁸https://developer.here.com/documentation/traffic/dev_guide/topics/what-is.html

⁹https://developer.here.com/documentation/traffic/topics_v6.1/resource-parameters-flow.html

maneira que os dados de tráfego e os arquivos *.osm* são organizados seguem padrões de identificação distintos. Ou seja, o identificador de uma via no padrão do HERE não pode ser convertido diretamente para o padrão do SUMO.

Isso também vale para as vias, pois elas possuem tamanhos distintos. No caso do *.osm*, cada interseção é tratada como uma aresta no grafo do cenário, considerando as intersecções como nós. Já no padrão do HERE, as diferentes arestas que compõem uma determinada via possuem tamanhos distintos sem considerar as intersecções. Para exemplificar algumas diferenças entre os dados, a Figura 2 apresenta os dados obtidos da cidade de Chicago por meio da Traffic API comparando-o com um cenário simulado. Observa-se que há uma maior concentração de dados das vias principais, e menos nas vias dos bairros com menor extensão.

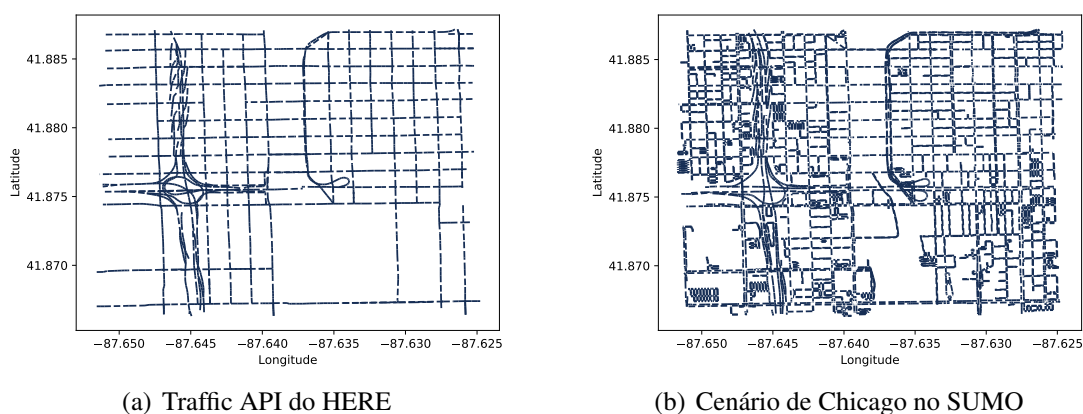


Figura 2. Comparação dos dados da API do Here e do SUMO.

Para mapear os dados é necessário considerar os pontos geográficos de cada aresta e verificar se elas estão sobrepostas. Além disso, como a ordem dos nós se refere a direção das arestas, é possível verificar a angulação para que não seja perdida a informação sobre a direção do tráfego. Por causa disso, dois problemas foram encontrados considerando a precisão das coordenadas geográficas, sendo elas: (i) incerteza em arestas paralelas, os pontos de origem e destino de cada aresta podem estar dispostos próximos mas não sobrepondo as arestas; e (ii) angulação entre as arestas, que pode apresentar uma pequena variação para mais ou para menos. Para resolver a incerteza, foi considerada uma distância aproximada do mapeamento real de cada aresta de 10 metros. Já para resolver a questão da angulação foi considerada uma diferença na angulação aceitável de 20 graus. O Algoritmo 1 apresenta como é feito o mapeamento dos dados de forma a identificar a relação entre os IDs de cada um.

Após o mapeamento dos IDs, é possível utilizar os dados do Here para visualizar o índice de tráfego em cada via do simulador. Esse resultado é apresentado na Figura 3. Para gerar os padrões de mobilidade a ferramenta *TrafficModeller*¹⁰ foi utilizada considerando o fluxo de tráfego coletado. Essa ferramenta requer como entrada apenas um arquivo no formato *.net.xml*. Então, é possível selecionar regiões de início e fim de rotas, quantidade de veículos, e horário de saída dos veículos podendo distribuí-los ao longo do tempo. Para decidir quais regiões serão utilizadas, é feita a análise dos fluxos obtidos ao longo do tempo.

Essa ferramenta escolhe pontos aleatórios para início e fim de cada rota dentro

¹⁰https://sumo.dlr.de/docs/Contributed/SUMO_Traffic_Modeler.html

Algoritmo 1: MAPEAMENTO DE TRÁFEGO

Entrada: *fluxoTrafego*, *grafoCenario*

```
1 início
2   mapeados ← { }
3   para cada fluxo ∈ fluxoTrafego faça
4     para cada aresta ∈ grafoCenario faça
5       \\ Verifica a diferença no ângulo em relação a um limite
           determinado
6       se diferencaAngulo(fluxo, aresta) < difAceitavel então
7         \\ Verifica se as arestas se intercedem
8         se verificaIntersecao(fluxo, aresta) então
9           \\ Faz o mapeamento entre o fluxo e a aresta
10          mapeados[aresta] = fluxo
11          fim
12        fim
13      fim
14    fim
15  retorna mapeados
16 fim
```

das regiões. Em outras palavras, algumas rotas podem não ser válidas considerando a inexistência de um caminho do início ao fim. Isso ocorre devido aos dados do *OpenStreetMap* serem gerados por *crowdsourcing* e conseqüentemente podem possuir erros, por exemplo, vias sem conexão com o restante do cenário. Utilizando a ferramenta *duarouter* as rotas são validadas de forma a garantir pelo menos um caminho válido em cada rota. Para simular foi necessário utilizar conjuntos distintos de rotas para verificar a validade da simulação. Sendo assim, o seguinte processo¹¹ foi iterado pela quantidade de conjuntos requeridos:

- Escolher aleatoriamente uma quantidade de rotas de cada região;
- Embaralhar a ordem das rotas para distribuir ao longo do tempo de maneira agrupada (de acordo com as regiões);
- Salvar o resultado de acordo com o padrão utilizado pelo SUMO alocando um ID para cada uma.

3.2. Identificação das Janelas Contextuais

O objetivo deste módulo é identificar com precisão as alterações que ocorrem nos dados utilizados. Ele é baseado em um trabalho anterior [Ladeira et al. 2019a], o qual possui diversas etapas, tais como: granularidade dos dados, descreve como é feita a filtragem e separação dos dados de acordo com diferentes variáveis temporais; pontuação da distribuição ao longo do tempo, descreve a utilização de uma função gaussiana para caracterizar a distribuição e então obter os pontos de crescimento e diminuição na ocorrência de eventos; identificação das janelas, fatiamento dos dados caracterizados de acordo com os “picos” e “vales” para obter as situações distintas; remoção de *outliers*, é aplicado um algoritmo de agrupamento para identificar quais eventos não representam a situação dentro de uma janela.

¹¹<https://github.com/lucaslzl/sumotripscleaner>

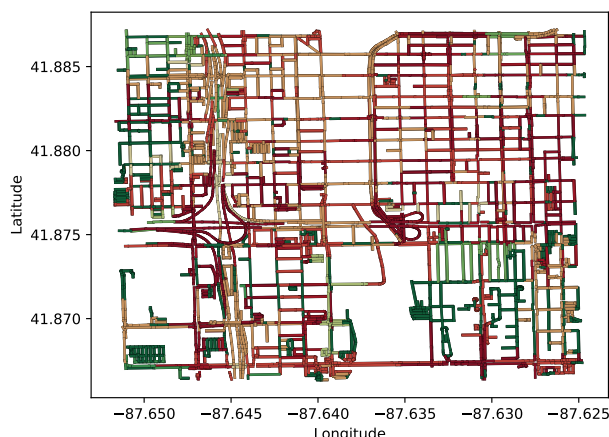


Figura 3. Mapeamento dos dados de tráfego as 14:10 de Segunda-feira.

Algumas melhorias foram aplicadas de forma a tornar o módulo mais robusto em relação a estratégia proposta anteriormente. A granularidade foi alterada considerando a não filtragem dos dados pela variável dos meses do ano, compreendendo apenas a separação pelos dias da semana, e os tipos e subtipos dos dados. Pela utilização de uma função gaussiana para descrever a distribuição dos eventos ao longo do tempo, é obtido um vetor com a pontuação em espaços de tempos previamente definidos. Ao observar esse vetor é possível observar um comportamento próximo de um sinal com algumas elevações e quedas abruptas. Portanto, foi adicionado um filtro entre as etapas de pontuação da distribuição ao longo do tempo e a identificação das janelas. Esse se refere ao filtro de wiener, comumente utilizado para processamento de sinais obtidos ao longo do tempo, com o intuito de filtrar ruído. Isso torna a caracterização da distribuição mais fácil para separar entre janelas distintas.

3.3. Mapeamento do Contexto

Esse módulo tem o objetivo de processar as janelas contextuais obtidas durante a sugestão das rotas e gerar distribuições de probabilidade para cada uma. De forma a evitar o processamento repetido, o resultado é armazenado enquanto a janela refletir a situação dos dados contextuais. Dessa maneira, os dados filtrados e separados nas janelas são mapeados no cenário considerando o horário da recomendação. A distribuição utilizada é uma *Kernel Density Estimation* com Kernel Gaussiano que permite que cada evento descreva uma área de impacto, somando a cada evento próximo para aumentar o risco. Seguindo a formulação de [Ledl 2004], uma KDE é descrita na Equação 1.

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K \frac{x - x_i}{h} \quad (1)$$

Na Equação 1, K é o kernel, portanto no caso deste trabalho é o Kernel Gaussiano. O valor h é um parâmetro de suavização, o qual deve ser estritamente maior que 0. O parâmetro x é o valor de referência, e por fim, x_i é um evento específico considerando o conjunto de eventos (x_1, x_2, \dots, x_n) . Na Figura 4, são apresentadas duas KDEs com os dados de criminalidade e acidentes envolvendo veículos da cidade de Chicago.

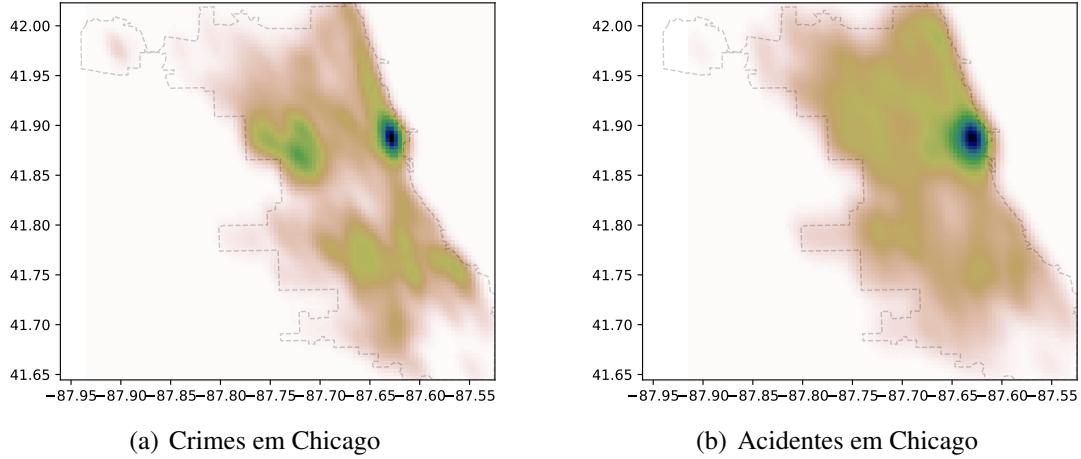


Figura 4. Kernel Density Estimations de diferentes dados contextuais

3.4. Personalização do Roteamento

Para atender diferentes perfis de usuários, foi modelado uma estratégia de soma ponderada que permite uma configuração direta do CERVA. Como resultado, o perfil do usuário se refere a pesos atribuídos, por exemplo, 70% para um determinado dado. O intuito é que o perfil informe quanto que cada tipo de dado contribui para o risco final. O peso deve ser definido para cada tipo de dado contextual utilizado com soma final de 100%. No entanto, primeiro, é necessário calcular a pontuação de cada tipo de dado contextual com o intervalo entre 0 a 1 para que todos tenham a mesma variação. Nesta etapa, os dados coletados em tempo real são utilizados para auxiliar a selecionar rotas. A pontuação do tráfego é obtida com base na velocidade máxima e média de cada estrada, variando entre 0 a 1, sendo 1 uma estrada congestionada. A Equação 2 apresenta a função de normalização utilizada para cada estrada r .

$$traffic_r = \frac{maxspeed_r - meanspeed_r}{maxspeed_r} \quad (2)$$

Então, um grafo ponderado de todas as vias é criado com base nas interseções (vértices) de cada via (arestas). A localização espacial de cada via é correlacionada com a pontuação p_i obtida por cada KDE em um determinado momento. Considerando a ocorrência de mais de um sub-tipo de dado, por exemplo o tipo do crime, apenas a maior pontuação é utilizada. Cada pontuação geral da via (r) é calculada de acordo com a estratégia da soma ponderada (Equação 3). A etapa de re-roteamento ocorre a cada t minutos, quando o veículo exige uma rota atualizada. Usando sua posição atual, o destino e o grafo das vias, o algoritmo Dijkstra é usado para encontrar o caminho mais curto (rota de pontuação geral mais baixa).

$$geral_r = \sum_{i=1}^n max(p_i) \cdot w_i \quad (3)$$

Um problema que pode ocorrer nesse cenário é a criação de novos congestionamentos por levar diversos veículos próximos para uma mesma via. Sendo assim, é necessário aplicar uma estratégia de balanceamento de carga para distribuição do tráfego.

Config	A	B	C	D	E	F	G
T/C/A	1.0/0.0/0.0	0.0/1.0/0.0	0.0/0.0/1.0	0.5/0.25/0.25	0.25/0.5/0.25	0.25/0.25/0.5	0.33/0.33/0.33
Config	H	I	J	K	L	M	Baseline
T/C/A	0.75/0.25/0.0	0.75/0.0/0.25	0.25/0.0/0.75	0.0/0.25/0.75	0.25/0.75/0.0	0.0/0.75/0.25	0.0/0.0/0.0

Tabela 1. Configurações de roteamento

A estratégia se resume em contar a quantidade de veículos re-roteados a cada via, considerando a capacidade (quantidade de faixas e tamanho da via) e aumentar o peso da via no grafo proporcionalmente ao seu percentual de carga. Tal estratégia é apresentada na Equação 4. Sendo tam o tamanho da via, qtd a quantidade de faixas, ζ o tamanho médio dos veículos (nesse caso 5), e $vload$ quantos veículos já foram roteados a uma determinada via r .

$$weight_r = \frac{vload}{\frac{tam \cdot qtd}{\zeta}} weight_r + weight_r \quad (4)$$

4. Avaliação do Sistema

4.1. Metodologia

A plataforma de simulação escolhida para validar o CERVA foi o SUMO (Simulador de Mobilidade Urbana)¹². O SUMO possui bibliotecas para implementar as chamadas de simulação em diferentes linguagens de programação. Embora algumas iniciativas tenham como objetivo criar um conjunto de dados contextuais como apresentado por Yu et al. [Yu et al. 2018], a iniciativa Open Data ainda é a maneira direta de obter dados da cidade. A cidade de Chicago foi escolhida com base na disponibilidade de dados abertos. Dados contextuais de crimes e acidentes de veículo foram obtidos e compreendem a localização espacial, o *timestamp* e o subtipo.

Uma área de $5km^2$ é utilizada nas simulações que compreende a área com maior incidência de eventos tanto para crimes como para acidentes. Um padrão de mobilidade foi gerado considerando o que foi apresentado na Subseção 3.1, e, então, partindo dele foram obtidas 20 variações. Portanto, a simulação foi executada 20 vezes para cada configuração. O horário da simulação tem início 12:00 de cada dia, e tem duração de 2 horas. Esse período foi escolhido por apresentar um grande número de registros (crimes e acidentes). Como é identificado o início e fim de cada situação (janela temporal), todas as mudanças que ocorrem nas situações durante esse período é considerado pelo sistema para sugerir novas rotas. As métricas obtidas na simulação referem-se ao somatório da pontuação calculada com base nas distribuições de probabilidade e o índice de tráfego. Foi calculada uma média considerando um intervalo de confiança de 95%.

Na avaliação do CERVA, foram considerados 14 perfis de roteamento distintos, de acordo com pesos para os dados contextuais. A Tabela 1 ilustra cada tipo de configuração e seu peso. Os pesos referem-se a cada tipo de dados contextuais, consequentemente, tráfego (T), crimes (C) e acidentes envolvendo veículos (A). Nas primeiras configurações, apenas os dados contextuais determinados são utilizados ignorando qualquer aspecto dos outros dois. O restante das configurações se referem as combinações de porcentagem de cada tipo de dado, desconsiderando o último que se refere ao *baseline*. O baseline se limita a não redirecionar os veículos, e usa apenas o padrão de mobilidade gerado inicialmente.

¹²<https://sumo.dlr.de/index.html>

Config	A			B			C		
T - C - A	0,806	16,889	28,083	2,205	8,746	14,204	2,218	8,707	14,176
Config	D			E			F		
T - C - A	1,704	8,726	14,36	1,967	8,695	14,198	1,998	8,693	14,145
Config	G			H			I		
T - C - A	1,916	8,681	14,251	1,192	9,519	16,271	1,384	9,122	14,902
Config	J			K			L		
T - C - A	2,008	8,719	14,171	2,197	8,697	14,139	1,845	8,59	14,453
Config	M			Baseline					
T - C - A	2,203	8,752	14,201	3,362	12,227	19,55			

Tabela 2. Resultados do roteamento

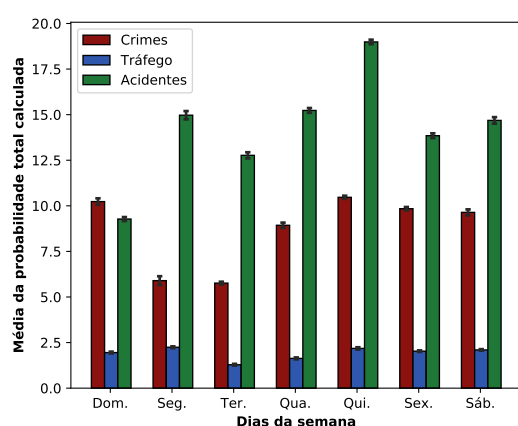
4.2. Resultados

Nesta subseção, os resultados das simulações são apresentados separados por configuração. Primeiro, as métricas contextuais são discutidas para verificar a execução correta do CERVA, alterando os dados contextuais considerados e verificando o impacto nessas métricas na tarefa de re-roteamento. Na Tabela 2, os resultados dessas métricas podem ser observados. Em negrito estão dispostos os melhores resultados de tráfego, crime, e acidentes entre todas as configurações utilizadas. Nesse caso, pode ser observado que o menor nível de tráfego foi obtido por considerar apenas o tráfego nas vias (Config A). Uma característica que difere o tráfego dos dados contextuais é a capacidade em mapear os estados futuros, podendo evitar congestionamentos por um gerenciamento eficiente. Isso não é observado no Baseline por não fazer gerenciamento algum de tráfego, tendo obtido o pior nível tráfego durante as simulações.

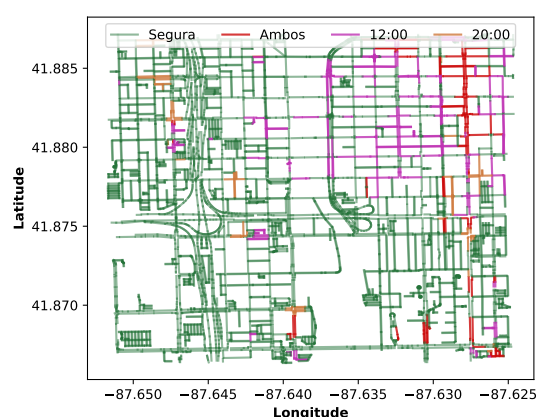
O nível de crime mais baixo foi obtido na configuração L, que considera a maior parte do peso para a criminalidade e uma pequena fração para o tráfego. Comparando a configuração B, que apenas considera a criminalidade, e a configuração L, foi observado o dobro de desvio padrão. Isso é uma consequência das mudanças na situação criminal que podem ocorrer durante o roteamento. Sendo assim, uma região que apresenta um nível de criminalidade baixo em um determinado momento, pode apresentar um aumento na criminalidade em outros, afetando todos os veículos dentro da região. Algo interessante ocorre para o nível de acidentes, sendo que, a configuração K apresenta o nível mais baixo. Observando o desvio padrão das outras configurações a maior parte é próximo do encontrado por K, caracterizando uma maior dificuldade em evitar regiões com risco de acidentes. Isso pode ser observado na Figura 4 (b).

As configurações podem ser observadas em grupos de acordo com a diferença entre os pesos, tais como: A, B e C (G1); D, E e F (G2); H e I (G3); J e K (G4), e L e M (G5). No caso de G2, é possível verificar que ao alterar os pesos o CERVA dá a preferência para o tipo de dado majoritário, isso sem prejudicar os outros. Com o aumento do peso para o tráfego em G3, é observado um aumento no nível dos outros. Em G4, é observada a maior diferença atingindo o menor nível de acidentes envolvendo veículos entre todas as configurações. O mesmo comportamento é observado em G5 para o nível de criminalidade. Considerando tudo o que foi verificado, é possível observar a capacidade do CERVA de sugerir rotas personalizadas para os usuários. Mesmo apenas utilizando nesse trabalho o tráfego, criminalidade, e acidentes envolvendo veículos, o CERVA pode ser estendido e utilizado com outros dados contextuais.

É possível também observar cada métrica entre os diferentes dias na semana. Para tal, é utilizada a configuração G que apresenta os mesmos pesos para todos os tipos de dados. Na Figura 5 (a), é possível observar a comparação entre as diferentes métricas. A terça-feira é o dia que apresenta os menores níveis para tráfego e crimes, como também é o segundo menor para acidentes. Já a quinta-feira apresenta os maiores níveis para todos os tipos de dados. Observando apenas os crimes, é possível ver uma diminuição durante a segunda-feira e terça-feira, e então ocorre um aumento que se mantém elevado até o resto da semana. Os níveis de acidentes apresentam um maior desvio padrão relativo aos demais tipos de dados. Nesse sentido, também ocorre uma maior variação deles entre dias seguidos. Por fim, é possível observar que a maneira que os padrões de mobilidade foram gerados permitem diferenciar o tráfego entre cada dia da semana.



(a) Comparação entre os dias da semana.



(b) Comparação das vias inseguras de acordo com horários distintos.

Figura 5. Resultados do sistema de acordo com os dias da semana e horas distintas.

Na Figura 5 (b), são apresentadas as vias do cenário, em verde as vias seguras, em rosa as inseguras as 12:00, em laranja as inseguras as 20:00, e por fim, as em vermelho quando em ambos horários a via é insegura. É possível observar que as 12:00 é mais complicado de evitar vias perigosas em comparação com as 20:00. Portanto, é possível verificar as mudanças durante a sugestão de rotas. No CERVA também é utilizada uma estratégia para identificar com maior granularidade as diferentes situações que ocorrem dentro de cada dia. Portanto, o CERVA tem a capacidade de identificar a situação para cada tipo de dado contextual e sugerir rotas melhores de acordo com o perfil do usuário.

5. Conclusões e Trabalhos Futuros

Este trabalho apresentou o CERVA, um serviço de sugestão de rotas para veículos personalizável e ciente de contexto. Para resumir seus benefícios de acordo com cada módulo, a identificação de janelas contextuais é capaz de identificar o comportamento de cada tipo de dado contextual ao longo do dia. Além disso, o CERVA supervisiona as alterações que ocorrem usando uma Função de Distribuição Gaussiana. O módulo de mapeamento de contexto caracteriza cada região da cidade com uma distribuição de probabilidade de acordo com cada janela contextual identificada. Por fim, a personalização do roteamento permite que os usuários atribuam suas preferências para evitar determinadas regiões contextuais. Além disso, esse módulo balanceia o tráfego para não criar novos pontos de congestionamento. Os resultados mostraram que o CERVA se adapta até a sensíveis

mudanças no peso de cada tipo de dados. Como trabalhos futuros foi identificada a possibilidade de expandir a quantidade de dados contextuais utilizados.

Agradecimentos. Os autores gostariam de agradecer o apoio financeiro concedido pelo CNPq (processo 167213/2018-0) para o bolsista de Mestrado Lucas Zanco Ladeira. Além desse, este trabalho foi parcialmente apoiado por dois projetos da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), sendo eles, o projeto GoodWeb (processo nº 2018/23011-1) e processo nº 2018/19639-5.

Referências

- Akabane, A. T., Immich, R., Madeira, E. R. M., and Villas, L. A. (2018). imob: An intelligent urban mobility management system based on vehicular social networks. In *2018 IEEE Vehicular Networking Conference (VNC)*, pages 1–8.
- Andresen, M. A. and Malleson, N. (2015). Intra-week spatial-temporal patterns of crime. *Crime Science*, 4(1):12.
- Cunha, F., Maia, G., Celes, C., Santos, B., Rettore, P., Campolina, A., Guidoni, D., Souza, F., Ramos, H., Villas, L., Mini, R., and Loureiro, A. (2017). Sistemas de transporte inteligentes: Conceitos, aplicações e desafios.
- Elsmore, S., Subastian, I. F., Salim, F. D., and Hamilton, M. (2014). Vdim: Vector-based diffusion and interpolation matrix for computing region-based crowdsourced ratings: Towards safe route selection for human navigation. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia, MUM '14*, pages 212–215, New York, NY, USA. ACM.
- Felson, M. and Boivin, R. (2015). Daily crime flows within a city. *Crime Science*, 4(1):31.
- Galbrun, E., Pelechris, K., and Terzi, E. (2016). Urban navigation beyond shortest route: The case of safe paths. *Information Systems*, 57:160 – 171.
- Ghorai, C. and Banerjee, I. (2017). A multi-objective data dissemination protocol for intelligent transportation systems. In *2017 IEEE 7th International Advance Computing Conference (IACC)*, pages 144–149.
- Gunantara, N. (2018). A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1):1502242.
- Jaimes, L. M. S., Ullah, K., and dos Santos Moreira, E. (2016). A secure commercial ads dissemination scheme for vehicular networks. In *2016 8th IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1–6.
- Keler, A. and Mazimpaka, J. D. (2016). Safety-aware routing for motorised tourists based on open data and vgi. *Journal of Location Based Services*, 10(1):64–77.
- Kim, J., Cha, M., and Sandholm, T. (2014). Socroutes: Safe routes based on tweet sentiments. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion*, pages 179–182, New York, NY, USA. ACM.
- Ladeira, L. Z., de Souza, A. M., Pereira, G., Silva, Thiago Henrique, M. F. S., and Villas, L. (2019a). Martini: Towards a mobile and variable time window identification for spatio-temporal data. In *IEEE Latin-American Conference on Communications*, pages 1–10. IEEE.
- Ladeira, L. Z., de Souza, A. M., Pereira, G., Silva, T. H., and Villas, L. (2019b). Serviço de sugestão de rotas seguras para veículos. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 608–621. SBC.

- Ledl, T. (2004). Kernel density estimation: Theory and application in discriminant analysis. *AUSTRIAN JOURNAL OF STATISTICS Volume Austrian Journal of Statistics*, 33:267–279.
- Liu, L., Chen, C., Qiu, T., Zhang, M., Li, S., and Zhou, B. (2018). A data dissemination scheme based on clustering and probabilistic broadcasting in vanets. *Vehicular Communications*, 13:78 – 88.
- Malleson, N. and Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population. *Crime Science*, 4(1):10.
- Newton, A. and Felson, M. (2015). Editorial: crime patterns in time and space: the dynamics of crime opportunities in urban areas. *Crime Science*, 4(1):11.
- Rezaei, M., Noori, H., Rahbari, D., and Nickray, M. (2017). Refocus: A hybrid fog-cloud based intelligent traffic re-routing system. In *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pages 0992–0998.
- Sahana, M., Hong, H., and Sajjad, H. (2018). Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: A study on kolkata urban agglomeration, india. *Science of The Total Environment*, 628-629:1557 – 1566.
- Santos, F. A., Akabane, A. T., Yokoyama, R. S., Loureiro, A. A. F., and Villas, L. A. (2016). A roadside unit-based localization scheme to improve positioning for vehicular networks. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–5.
- Santos, F. A., Rodrigues, D. O., Silva, T. H., Loureiro, A. A. F., Pazzi, R. W., and Villas, L. A. (2018). Context-aware vehicle route recommendation platform: Exploring open and crowdsourced data. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–7.
- Secron, T. M., da Silva, E. R., de Farias, C. M., and Cruz, T. (2016). Sigaciente: Uma ferramenta para inferência do trânsito e de rotas seguras baseada em dados sociais. In *ERSI'2016*, pages 58–65.
- Shah, S., Bao, F., Lu, C.-T., and Chen, I.-R. (2011). Crowdsafe: Crowd sourcing of crime incidents and safe routing on mobile devices. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '11*, pages 521–524, New York, NY, USA. ACM.
- Souza, A., Braun, T., and Villas, L. (2018). Efficient context-aware vehicular traffic re-routing based on pareto-optimality: A safe-fast use case. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2905–2910.
- Souza, A. M. d., Braun, T., Botega, L. C., Cabral, R., Garcia, I. C., and Villas, L. A. (2019). Better safe than sorry: a vehicular traffic re-routing based on traffic conditions and public safety issues. *Journal of Internet Services and Applications*, 10.
- Utamima, A. and Djunaidy, A. (2017). Be-safe travel, a web-based geographic application to explore safe-route in an area. *AIP Conference Proceedings*, 1867(1):020023.
- Yu, T.-Y., Zhu, X., and Maheswaran, M. (2018). *Vehicular Crowdsensing for Smart Cities*, pages 175–204. Springer International Publishing, Cham.
- Zheng, Y., Zhang, Y., and Li, L. (2016). Reliable path planning for bus networks considering travel time uncertainty. *IEEE Intelligent Transportation Systems Magazine*, 8(1):35–50.