

Chove lá fora, alerta aqui dentro: Uma análise de Alertas de Problemas de Trânsito em Sensoriamentos Participativos

Iran F. Ribeiro¹, Vitor F. Calmon¹, Thiago H. Silva², Vinícius F. S. Mota¹

¹Departamento de Informática – Universidade Federal do Espírito Santo
Vitória – Brasil

{iran.ribeiro, vitor.calmon}@edu.ufes.br, {vinicius.mota}@inf.ufes.br

²Departamento de Informática – Universidade Tecnológica Federal do Paraná
Curitiba – Brasil

thiagoh@utfpr.edu.br

Abstract. *The urban population has increased rapidly in recent years, and problems involving the quality of life of people in urban centers have also grown with it. In this sense, the evolution of Information and Communication Technologies (ICTs) allows crowdsensing-based services to be an essential tool to improve people's quality of life by providing valuable data to public agents who manage such centers. This work proposes the analysis of a event (problems) dataset reported by users of a crowdsensing app in Vitória-ES, considering three questions: the duration of the reported events, monthly variation of events, and the correlation between rainfall in the city and events reported by users. Among other results, we observed that most of the studied events have an unclear clear seasonality pattern and could be influenced by climatic factors. We also observed positive (low and moderate) correlations between rainfall and events, especially when analyzing individual months. This result can open new ways of investigation for, for example, new services for forecasting traffic problems.*

Resumo. *A população urbana tem crescido rapidamente nos últimos anos e com ela aumentam também os problemas envolvendo a qualidade de vida das pessoas nos centros urbanos. Nesse sentido, a evolução das Tecnologias da Informação e Comunicação (TICs), permite que serviços baseados em crowdsensing sejam uma ferramenta essencial para melhorar a qualidade de vida das pessoas fornecendo dados valiosos para agentes públicos que gerenciam tais centros. Este trabalho propõe a análise de uma base de dados de eventos (problemas) de trânsito reportados por usuários em um aplicativo de sensoriamento participativo de tráfego na cidade de Vitória-ES considerando três questões: o tempo de duração dos eventos reportados, variação mensal dos eventos, e a correlação entre as chuvas na cidade e os eventos reportados pelos usuários. Dentre os resultados, foi observado que a maioria dos eventos estudados não apresentam um padrão claro de sazonalidade, o que poderiam estar sendo influenciados por fatores climáticos. Observamos ainda correlações positivas (baixas e moderadas) entre chuvas e eventos, principalmente ao analisar meses individualmente. Esse resultado pode abrir novos caminhos de investigação para, por exemplo, novos serviços de previsão de problemas de trânsito.*

1. Introdução

O aumento da porcentagem da população que vive em áreas urbanas e, consequentemente, dos problemas causados por tal aumento, incentiva a utilização de tecnologias da informação e comunicação como uma forma de minimizar o impacto dessas mudanças na vida das pessoas [Ismagilova et al. 2019]. Pesquisas nesse sentido visam, por exemplo, monitorar a qualidade do ar [SM et al. 2019, Gulia et al. 2020], água [Pule et al. 2017, Ighalo and Adeniyi 2020] e oferecer informações relevantes para gestores de centros urbanos, como dados relacionados à mobilidade urbana [Solmaz et al. 2019, Joo et al. 2020]. Enquanto a instalação de infraestrutura fixa para monitoramento da cidade pode ser caro e de difícil manutenção, o sensoriamento participativo, também chamado de *crowdsensing*, pode permitir uma coleta de dados distribuída e mais eficiente.

Em especial, em uma época em que as pessoas estão cada vez mais conectadas, dados provenientes de serviços de *crowdsensing* são fundamentais para fornecer soluções aos problemas do meio urbano. Especificamente em relação à mobilidade urbana, diversos estudos investigaram como esses dados podem auxiliar agentes públicos que monitoram e gerenciam o trânsito das cidades. Tais estudos vão desde a detecção e monitoramento de acidentes [dos Santos et al. 2017, Lin and Li 2020], detecção de anomalias [Thomé et al. 2020], estudo dos padrões de incidentes [Silva et al. 2013], detecção de incidentes [Lenkei 2018] e se baseiam principalmente em combinar fontes diferentes de dados para tratar de um problema específico. Assim, embora abordado brevemente em [dos Santos et al. 2017, Silva et al. 2013], os estudos não avaliam as possíveis causas e influência de eventos externos nos dados fornecidos pelos usuários de *crowdsensing*.

Dessa forma, este trabalho propõe a análise de fatores externos, particularmente chuvas, em uma base de dados de um aplicativo de navegação baseado em *crowdsensing* na cidade de Vitória, no Espírito Santo. Cada registro na base de dados corresponde a um evento que indique algum problema em uma rua/avenida da cidade, que é reportado por um usuário do aplicativo. Todos esses eventos estudados poderiam ter relações com chuvas. A análise é feita com o objetivo de entender como os usuários interagem com o aplicativo e visa responder à 3 perguntas sobre os dados: i) Qual a duração média dos eventos reportados? ii) Existe algum padrão ou sazonalidade de número de eventos reportados? iii) Utilizando dados meteorológicos de fontes oficiais, é possível identificar influência das chuvas sobre os eventos reportados? Desta forma, o trabalho visa caracterizar os eventos que mais causaram transtornos na cidade, isto é, quanto tempo um evento incomodou, se o transtorno é recorrente, e os efeitos de eventos naturais, como a chuva.

Nossos resultados indicam que a maioria dos eventos tem um tempo de duração menor que aproximadamente 10 dias, considerando todas as ruas da cidade e não apresentam sazonalidade nos meses observados. Além disso, a partir de dados do Instituto Nacional de Meteorologia (INMET), encontramos evidências de que existe uma correlação positiva (baixa/moderada) de chuvas nos eventos estudados. Esse tipo de indicação pode abrir novas frentes de estudo visando o desenvolvimento de novas aplicações e serviços que objetivando melhorar a previsão de problemas causados por chuvas nas cidades.

O restante deste artigo está organizado como segue. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve as bases de dados e metodologia utilizadas. A Seção 4 discute os resultados obtidos a partir das análises feitas. Por fim, as considerações finais e trabalhos futuros são apresentadas na Seção 5.

2. Trabalhos Relacionados

Utilizando dados do *Waze*, [Silva et al. 2013] realiza um dos primeiros estudos sobre os padrões de incidentes reportados pelos usuários, mostrando a potencialidade desse tipo de dado em novas aplicações para o entendimento do trânsito. Ainda usando dados do *Waze*, [Lenkei 2018] faz um extenso estudo sobre a detecção de incidentes incomuns na cidade de Estocolmo, na Suécia. O autor realiza diversas análises sobre os dados, como análises descritivas e qualitativas, da confiabilidade dos dados fornecidos pelo *Waze* em relação aos incidentes reportados, análise de incidentes relacionados (influência de um incidente em outros), análise do impacto que os incidentes causam no trânsito. Por fim, a validação da detecção dos dados do *Waze* foi feita utilizando os dados de agências oficiais suecas por meio de regressão logística. Em comparação com os dados oficiais, os resultados mostram que é possível identificar 4.5 vezes mais eventos pelo *Waze*, com $\approx 27.7\%$ dos incidentes sendo identificados mais rápido.

Buscando mostrar a importância da utilização de dados de fontes heterogêneas para análises de fenômenos urbanos, [dos Santos et al. 2017] analisa dados de acidentes de trânsito fornecidos pelo *Waze* e de fontes oficiais da cidade de Belo Horizonte, MG. O processo de junção dos dados heterogêneos mostrou que é possível observar um aumento em cerca de 100.5% no número de acidentes reportados utilizando os dados do *Waze*, com apenas 7% dos acidentes sendo reportados em ambas as fontes. Isso reforça a ideia de que a combinação das fontes de dados distintas pode fornecer informações complementares sobre os acidentes na cidade. Além disso, os autores analisam brevemente a influência da variação da pluviosidade tempo na ocorrência dos acidentes reportados no *Waze*.

Utilizando dados de diferentes fontes, Thomé et al. [Thomé et al. 2020] propõe um arcabouço para a detecção e geração de alertas de anomalias na cidade de Vitória, no Espírito Santo. O arcabouço utiliza dados de um aplicativo de sensoriamento participativo (do inglês, *crowdsensing*) fornecidos pela prefeitura de Vitória. Os dados contêm informações como velocidade média nas vias congestionadas, e as anomalias detectadas em uma via consistem de congestionamentos não condizentes com o esperado para tal via em um dado horário. A fim de validar as anomalias detectadas, os autores utilizam dados do *Twitter* como segunda fonte do arcabouço. Nesse sentido, os alertas gerados podem auxiliar as tomadas de decisão de órgãos responsáveis pelo trânsito da cidade.

Lin e Li [Lin and Li 2020] investigam métodos para prever o comportamento do trânsito após o acontecimento de acidentes. A partir de dados de um aplicativo de navegação cujo nome não foi mencionado, os autores dividem o tráfego em 4 níveis de intensidade, de forma a prever a duração de cada nível ao longo do tempo após os acidentes. A predição é feita com algoritmos de aprendizado de máquina, como as *Random Forest*, *Support Vector Machine* e Redes Neurais. A partir dos experimentos, percebe-se que as redes neurais tem um melhor desempenho que os outros dois modelos na maioria dos casos. Entretanto, todos os modelos tendem a apresentar resultados melhores à medida que dados atualizados são fornecidos aos modelos.

Como pode ser visto, vários trabalhos na literatura combinam dados externos com dados de *crowdsensing* como forma de aumentar a confiabilidade dos estudos feitos ou obter dados com informações que se complementem. Entretanto, dado que a participação dos usuários é fundamental para dados de *crowdsensing*, abordamos tais dados de forma distinta dos demais estudos. Assim, neste trabalho investigamos

como eventos externos influenciam na interação de pessoas com um aplicativo de *crowdsensing*. Especificamente como estudo de caso, analisamos a influência de chuvas possui nos eventos reportados por usuários de um aplicativo de navegação entre na cidade de Vitória-ES. Além disso, para auxiliar nosso estudo e entendimento dos dados utilizados, discutimos os tempos de duração dos eventos reportados e investigamos como eles estão correlacionados entre si durante o período observado.

3. Bases de dados

Nesta seção é apresentada uma breve descrição das bases de dados utilizadas neste trabalho: Uma base de dados de eventos de trânsito da cidade de Vitória-ES, denominada de EventosVitória, cedidos pela prefeitura da cidade; e os dados de pluviômetros oficiais do Instituto Nacional de Meteorologia (INMET) também da cidade Vitória-ES. Além disso, descrevemos as informações selecionadas das respectivas bases e o processamento aplicado aos dados obtidos.

3.1. Base de dados EventosVitória

Esta é uma base de dados de um aplicativo de navegação baseada em *crowdsensing* composta por diversas informações sobre o trânsito da região da Grande Vitória, no Espírito Santo. Essa base de dados é resultado de uma parceria entre a prefeitura de Vitória e o aplicativo de sensoriamento participativo tráfego em tempo real *Waze* para dispositivos móveis. No aplicativo, os usuários reportam eventos atípicos nas vias, como acidentes, alagamentos ou engarrafamentos.

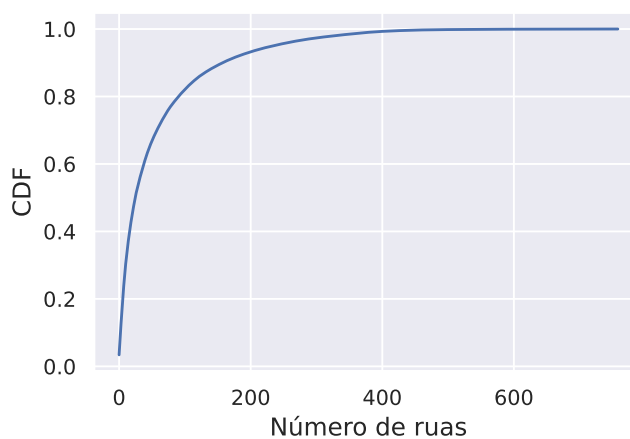


Figura 1. CDF do número de eventos reportados por rua.

Através de uma API fornecida pela desenvolvedora do aplicativo, a prefeitura de Vitória obtém acesso as informações sumarizadas e anonimizadas do aplicativo. Por sua vez, a prefeitura fornece uma API que permite coletar dados de três bases com informações distintas, sumarizadas a cada 5 minutos: eventos (eventos informados pelos usuários), congestionamentos e irregularidades (ambos mensurados pelo aplicativo). Para este trabalho utilizamos a primeira base de dados. Ressalta-se que esta é uma base de dados esparsa, pois utiliza informações inseridas pelos usuários, de modo que alguns horários não apresentam registros de eventos.

Assim, os dados utilizados estão compreendidos entre 17 de Março de 2019 e 02 de Dezembro de 2019, com 4 tipos e 24 subtipos de eventos que, por sua vez, foram reportados em 759 ruas. Para as análises definidas neste estudo, delimitamos o intervalo entre Março e Novembro. A Figura 1 apresenta a CDF do número de eventos reportados por rua na cidade de Vitória. Nessa figura nota-se que a maioria dos eventos acontece em poucas ruas, com $\approx 90\%$ dos eventos sendo reportados em $\approx 20\%$ das ruas. Por fim, a Tabela 1 mostra a proporção dos eventos reportados, evidenciando que o evento HAZARD_ON_ROAD_POT_HOLE, subtipo de WEATHERHAZARD, corresponde a 46.453% dos registros observados. Em outras palavras, entende-se a maioria dos eventos reportados está relacionado à eventos meteorológicos na cidade.

Tabela 1. Proporção dos tipos e subtipos de eventos reportados pelos usuários.

Tipo	Subtipo	%
WEATHERHAZARD	HAZARD_ON_ROAD_POT_HOLE	46.453
	HAZARD_ON_ROAD_CONSTRUCTION	2.737
	HAZARD_ON_ROAD_CAR_STOPPED	1.518
	HAZARD_WEATHER_FLOOD	0.507
	HAZARD_ON_SHOULDER_CAR_STOPPED	0.447
	HAZARD_ON_ROAD_OBJECT	0.173
	HAZARD_ON_ROAD_TRAFFIC_LIGHT_FAULT	0.150
	HAZARD_ON_SHOULDER_MISSING_SIGN	0.138
	Outros	0.115
JAM	JAM_HEAVY_TRAFFIC	11.320
	JAM_STAND_STILL_TRAFFIC	7.235
	JAM_MODERATE_TRAFFIC	3.373
ROAD_CLOSED	ROAD_CLOSED_EVENT	24.545
	ROAD_CLOSED_CONSTRUCTION	0.065
	ROAD_CLOSED_HAZARD	0.002
ACCIDENT	ACCIDENT_MINOR	0.798
	ACCIDENT_MAJOR	0.425

3.2. Base de dados pluviais

Os dados das chuvas em Vitória foram coletados do Instituto Nacional de Meteorologia (INMET)¹ através da estação convencional VITORIA (83648), localizada ao sul da cidade de Vitória, no Espírito Santo (Figura 2). Os dados foram coletados entre os meses de Março e Novembro de 2019, para corresponder ao intervalo da base EventosVitória, e possuem informações como temperatura, umidade, pressão, velocidade e direção do vento e precipitação, em milímetros.

Como desejamos avaliar a influência das chuvas nos eventos reportados, utilizamos as informações referentes à precipitação por dia no intervalo definido. A Figura 3 apresenta a média e desvio padrão do índice pluviométrico de cada dia do mês no período considerado. Nota-se que, em média, o índice é maior no meio de cada mês, e menor no início e fim. Embora outros fatores possam influenciar a quantidade de eventos reportados (ventos muito fortes podem derrubar alguma estrutura em uma via, por exemplo),

¹<https://portal.inmet.gov.br/>



Figura 2. Localização da estação convencional em Vitória, no Espírito Santo (os pontos roxos representam estações pluviométricas automáticas).

supomos que as chuvas possuem um impacto maior no trânsito e, conseqüentemente, nos eventos reportados pelos usuários do aplicativo de navegação.

Por fim, como há mais de um valor referentes às chuvas no dia, é preciso que seja obtido um valor único referente ao dia da coleta. Além disso, a estação pode apresentar falhas durante a mensuração resultando em um dado faltante para um determinado dia. O tratamento dessas questões e a junção dos dois *datasets* será discutido na próxima Seção.

3.3. Processamento dos dados

Na base EventosVitória, o primeiro processamento aplicado foi selecionar apenas as colunas com utilidade para o estudo atual. Tais colunas são *eventDate*, *type*, *subtype*, que representam, respectivamente, a data do acontecimento, o tipo e subtipo do evento. Em seguida, removemos os dados referentes à Avenida Leitão da Silva, pois todos eventos reportados nessa rua eram do tipo ROAD_CLOSED e provavelmente motivados pelo fato de a rua estar em reforma durante o período de observação dos dados.

Removemos, também, os registros com alguma coluna faltante nos dados. Todas as colunas são utilizadas para realizar algum tipo de seleção ou agrupamento dos dados e, nesse caso, não é possível inferir os dados faltantes a partir dos valores existentes. Durante a análise exploratória dos dados notamos que era comum o mesmo evento aparecer várias vezes em sequência. Como os valores na coluna *eventDate* eram diferentes, infere-se que não se tratava de dados faltantes e, sim, vários usuários reportando o mesmo evento. Nesse caso, julgamos que a remoção desses dados dificulta identificar eventos que causaram mais interação dos usuários com o aplicativo. Por fim, corrigimos as datas da base de dados, que são registrados em um fuso horário diferente do Espírito Santo.

Os dados pluviais passaram por dois processamentos principais: i) tratamento dos dados faltantes e ii) obtenção do índice pluviométrico diário. Dias com dados faltantes foram substituídos com o valor 0. Dado que um evento pode ser registrado à qualquer hora por um usuário, utilizamos o índice máximo de pluviosidade em milímetros de um dia como o valor referente àquele dia.

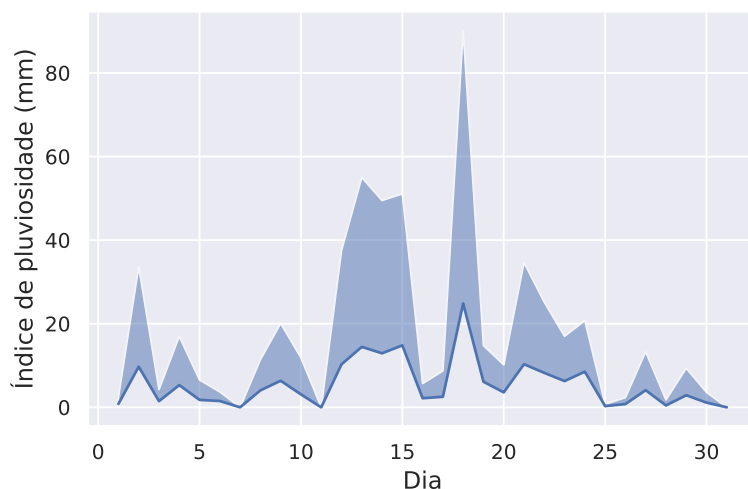


Figura 3. Índice pluviométrico em milímetros de cada dia do mês. A parte sombreada representa o desvio padrão das chuvas

Para a junção dos dois conjuntos de dados, identificou-se quando ambos têm início e fim simultâneos. Nesse sentido, a base EventosVitória, possuía eventos com alguns dias de diferença em que aparecem na base de dados. Nesses casos, identificamos as datas de início e fim dos eventos, de forma que todos estivessem no mesmo intervalo. O intervalo da base de dados resultante foi de 17 de Março de 2019 à 30 de Novembro de 2019.

3.4. Duração e Séries Temporais dos Eventos

Para avaliar como os usuários reportam os eventos na base de dados e como as chuvas os influenciam, calculamos 3 métricas: i) A duração de cada tipo de evento; ii) O número de eventos reportados (por dia, agrupados por mês) e; iii) A correlação entre os dados das chuvas e das séries temporais dos eventos reportados.

Para obter i) calculamos a duração dos eventos a partir do Algoritmo 1. O funcionamento do algoritmo consiste em, basicamente, percorrer todos os eventos presentes em um *dataset* e armazenar as datas inicial (linha 13) e final do evento em um dicionário. Como um evento pode ocorrer mais de uma vez e em datas distintas, a data final é atualizada à medida que o *dataset* é percorrido (linha 11). Nesse sentido, caso o evento seja visto com um intervalo de tempo maior que t em relação ao registro anterior, considera-se que o evento trata de uma ocorrência distinta e uma nova entrada é adicionada ao dicionário (linhas 7 a 9). Em todo caso, apenas a data final do registro vigente é atualizada. Por fim, as linhas 15 e 16 contabilizam a duração de cada evento presente no dicionário.

De maneira similar, para a obtenção do resultado ii) contabilizamos o número de vezes que cada evento foi reportado pelos usuários em um determinado intervalo de tempo, resultando em uma série temporal do respectivo evento. Essa contabilização é feita utilizando o Algoritmo 2. Dado um *dataset* D , o algoritmo salva o número de eventos reportados em uma janela de tempo t definida como parâmetro (linhas de 5 a 7). Essa contabilização permite analisar a evolução da colaboração dos usuários ao longo do tempo, bem como avaliar a correlação entre os dados das chuvas e os eventos reportados.

Por fim, para o resultado iii), as séries temporais dos eventos precisam ter o mesmo

Algoritmo 1: Calcula a duração dos eventos

Entrada: Dataset *Eventos*, Timedelta *t*

Saída: Dicionário com a duração dos eventos *duracao_evento*

```
1 dict_e ← {};  
2 duracao_evento ← {};  
3 foreach e ∈ Eventos do  
4   dt_atual ← e.eventDate;  
5   if e já foi visto then  
6     dt_final ← DtFinal(dict_e[e]);  
7     if (dt_atual - dt_final > t) then  
8       dt_inicial ← dt_final ← dt_atual;  
9       InsereNovoRegistro(dict_e[e], dt_inicial, dt_final);  
10    else  
11      AtualizaDataFinal(dict_e[e], dt_atual);  
12  else  
13    dt_inicial ← dt_final ← dt_atual;  
14    InsereNovoRegistro(dict_e[e], dt_inicial, dt_final);  
15 foreach e ∈ Eventos do  
16   duracao_evento[e] ← DtFinal(dict_e[e]) - DtInicial(dict_e[e]);
```

número de registros que a base de dados das chuvas. Para garantir isso, definimos o parâmetro *t* no Algoritmo 2 como 1 dia e utilizamos registros das duas bases de dados que estão no mesmo intervalo. A correlação utilizada é dada pela Equação 1

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (1)$$

onde *X* e *Y* são duas variáveis tal que $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$, $Cov(X, Y)$ calcula a covariância entre *X* e *Y*, σ_X é o desvio padrão da variável *X* e σ_Y o desvio padrão da variável *Y*. Essa equação é conhecida como Correlação de Pearson.

Algoritmo 2: Obtém a serie temporal de um *dataset* de eventos

Entrada: Dataset *D*, Timedelta *t*

Saída: Série temporal dos eventos *S*

```
1 di ← DtInicial(D);  
2 df ← DtFinal(D);  
3 S ← [];  
4 while di < df do  
5   proxData ← di + t;  
6   n_eventos ← número de eventos em D entre di e proxData;  
7   insere n_eventos em S;  
8   di ← proxData;
```

4. Resultados e Discussão

Nesta seção apresentamos as caracterizações feitas a partir dos dados e a análise da influência que as chuvas possuem nos eventos reportados. Além disso, discutimos a relação que os eventos reportados possuem entre si.

4.1. Duração dos eventos

Considerando um intervalo máximo de 5 dias, a Figura 4 mostra o diagrama de caixa para os 4 tipos de eventos na base de dados EventosVitória. Nota-se que JAM (engarrafamento) é o evento com maior duração entre os reportados, com uma mediana de aproximadamente 11 dias e duração máxima de pouco mais de 92 dias. Isso significa que 50% dos eventos desse tipo foram reportados durante 10 dias seguidos na cidade e, em alguns casos, e podendo a chegar à 92 dias. Além disso, nota-se que a maioria dos eventos tem uma duração inferior à 11 dias. Como consideramos todas as ruas da cidade, é improvável que um engarrafamento não ocorra em um dia. Isso justifica, por exemplo o fato de JAM ter maior duração e ser apenas o segundo evento mais reportado, como mostrado anteriormente na Tabela 1.

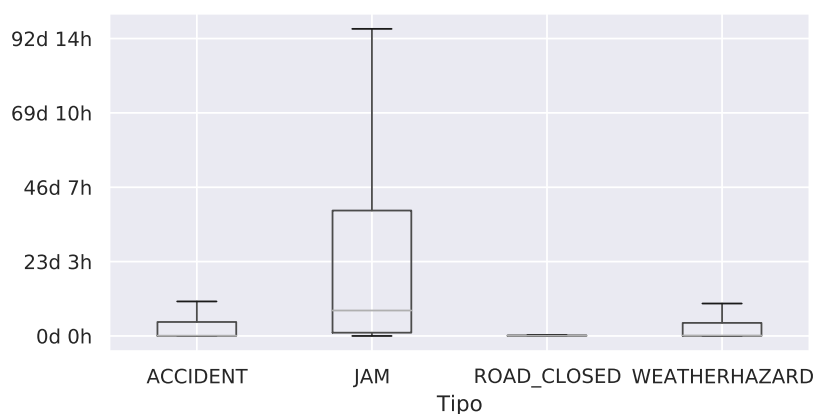


Figura 4. Diagrama de caixa da duração dos eventos

A figura mostra que os eventos ACCIDENT e WEATHERHAZARD têm, no geral, duração parecidas, com máxima de aproximadamente 11 dias e uma mediana baixa, de cerca de algumas horas. Assim como no evento JAM, esses eventos podem ocorrer em qualquer rua da cidade. Em outras palavras, os 11 dias em que os eventos ocorreram significam que, durante 11 dias, os eventos continuaram ocorrendo pela cidade e não necessariamente na mesma rua. Por fim, o evento ROAD_CLOSED apresenta a menor duração entre os eventos reportados, com eventos que duram apenas algumas horas.

4.2. Eventos reportados por dia

Para avaliar a variação dos 4 tipos de eventos reportados durante os dias entre Março e Novembro de 2019, computamos o número de eventos reportados em intervalos de 1 dia durante cada mês, na Figura 5. A figura mostra que, no geral, poucos eventos do tipo ACCIDENT (acidente) são reportados ao longo dos meses. Os outros eventos, por outro lado, ocorrem praticamente em todos os dias de cada mês. Além disso, os eventos oscilam sem nenhum padrão aparente, com exceção do JAM, que apresenta uma sequência de

valores muito baixos e muito altos. No primeiro caso, provavelmente são os finais de semana e, no segundo, picos que ocorrem nos dias da semana, com alguns dias tendo um pico maior que outros.

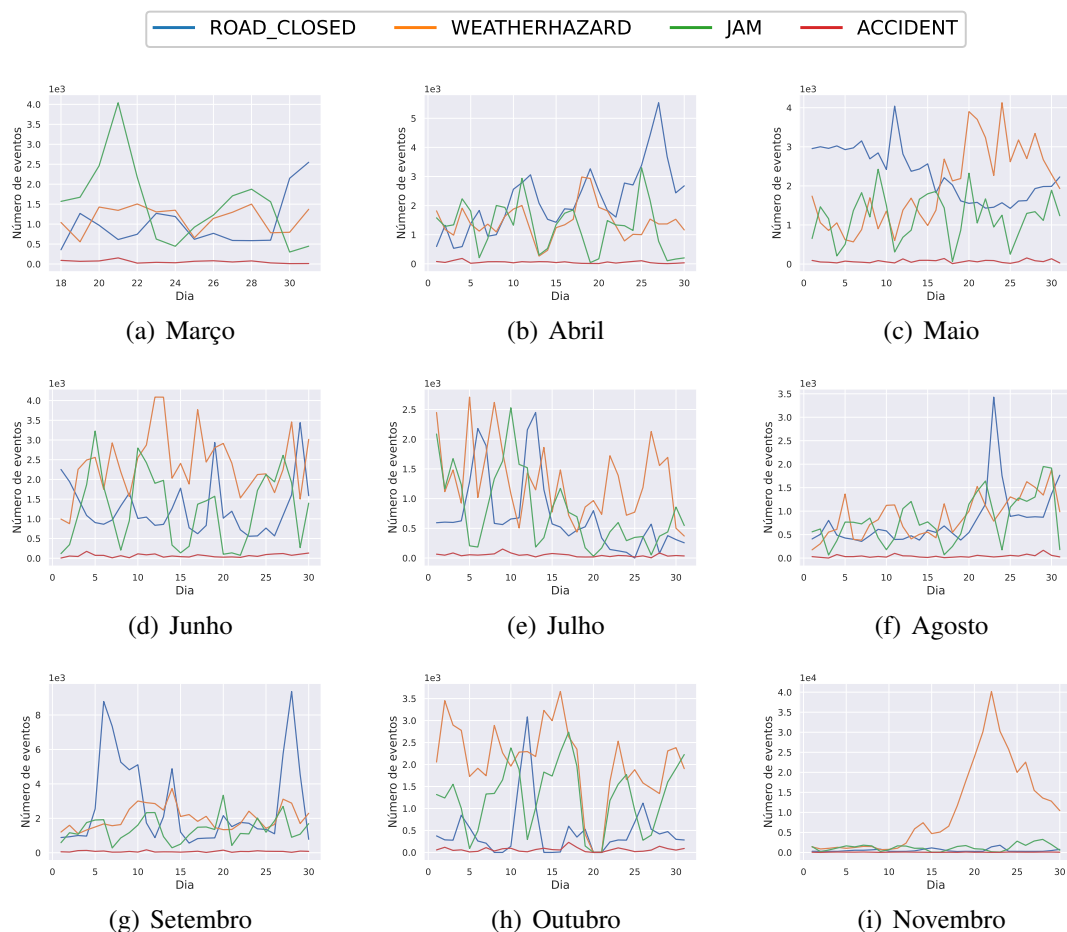


Figura 5. Eventos reportados por dia, entre Março e Novembro

Nas Figuras 5(b) (Abril) e 5(f) (Agosto) nota-se que ocorre um pico de registros do evento ROAD_CLOSED no final do mês, de ≈ 6000 e ≈ 3500 eventos, respectivamente. Outros dois picos de ROAD_CLOSED podem ser vistos em Setembro, na Figura 5(g), com aproximadamente ≈ 9500 eventos reportados. Por fim, em Novembro, na Figura 5(i), nota-se um pico significativo no número de eventos do tipo WEATHERHAZARD, com ≈ 39000 eventos. Os outros eventos, por outro lado, são reportados menos de 5000 vezes. Dado o grande volume de eventos do tipo WEATHERHAZARD, supõe-se que variações no índice de pluviosidade na cidade tenham acontecido e justifiquem ao menos parte das causas dos eventos.

A Figura 6 apresenta uma visão geral do resultado anterior, com o número de eventos reportados durante todo o mês. Fica evidente, por exemplo, a grande diferença entre os números de eventos do tipo ACCIDENT e dos demais. É interessante notar também que o evento JAM aparenta ter aproximadamente o mesmo número de registros durante os meses. E, como visto anteriormente, o evento WEATHERHAZARD apresenta um pico enorme de registros em novembro, bem acima dos outros meses.

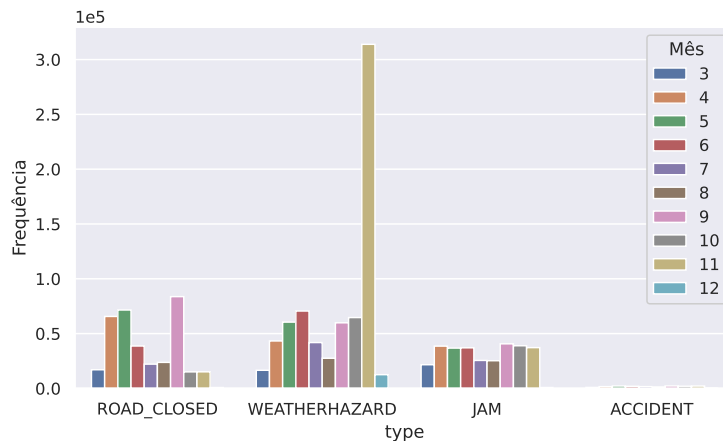


Figura 6. Frequência de eventos reportados por mês

4.3. Influência das chuvas nos eventos

Para avaliar a influência do clima, especificamente das chuvas, na interação dos usuários com o aplicativo de navegação, obtemos a correlação da base de dados das chuvas obtida do INMET com os 4 tipos de eventos reportados pelos usuários entre Março e Dezembro de 2019, na Figura 7. Como esperado, a figura evidencia que as chuvas têm maior correlação com o evento WEATHERHAZARD (0, 25). Quando olhamos os outros eventos, por outro lado, notamos que possuem correlação negativa baixa, praticamente inexistente, de $-0,1$ para JAMS e ACCIDENT e de $\approx -0,009$ para ROAD_CLOSED. Além disso, os eventos JAMS e ACCIDENT possuem uma correlação significativa de 0, 59. Essa correlação muito provavelmente vem do fato de que a ocorrência de um acidente causa congestionamentos no trânsito, como discutido em [dos Santos et al. 2017]. Esse é um resultado curioso porque, embora uma mínima correlação entre os dois eventos seja esperada, os resultados anteriores mostram que os acidentes correspondem a apenas $\approx 1,2\%$ dos eventos (ver Tabela 1).

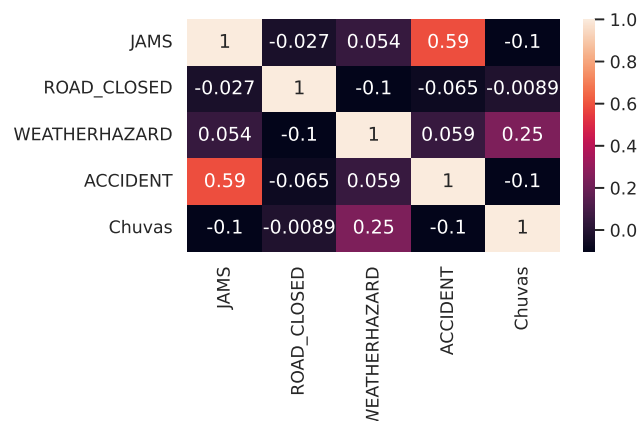


Figura 7. Correlação entre as chuvas e os eventos reportados.

Resultados anteriores mostram que há um grande volume de eventos do tipo WEATHERHAZARD no mês de Novembro. Porém, como as chuvas podem ocorrer com

mais ou menos frequência em alguns meses, calculamos a correlação das chuvas em cada mês entre Março e Novembro, que pode ser vista na Figura 8. De fato, nota-se que a correlação das chuvas com o evento WEATHERHAZARD é maior em Novembro (0,341), seguido de Setembro (0,292) e Outubro (0,234). É interessante notar também a correlação entre as chuvas e outros eventos como JAMS (0,359) em Setembro e ROAD_CLOSED (0,313) em Julho.

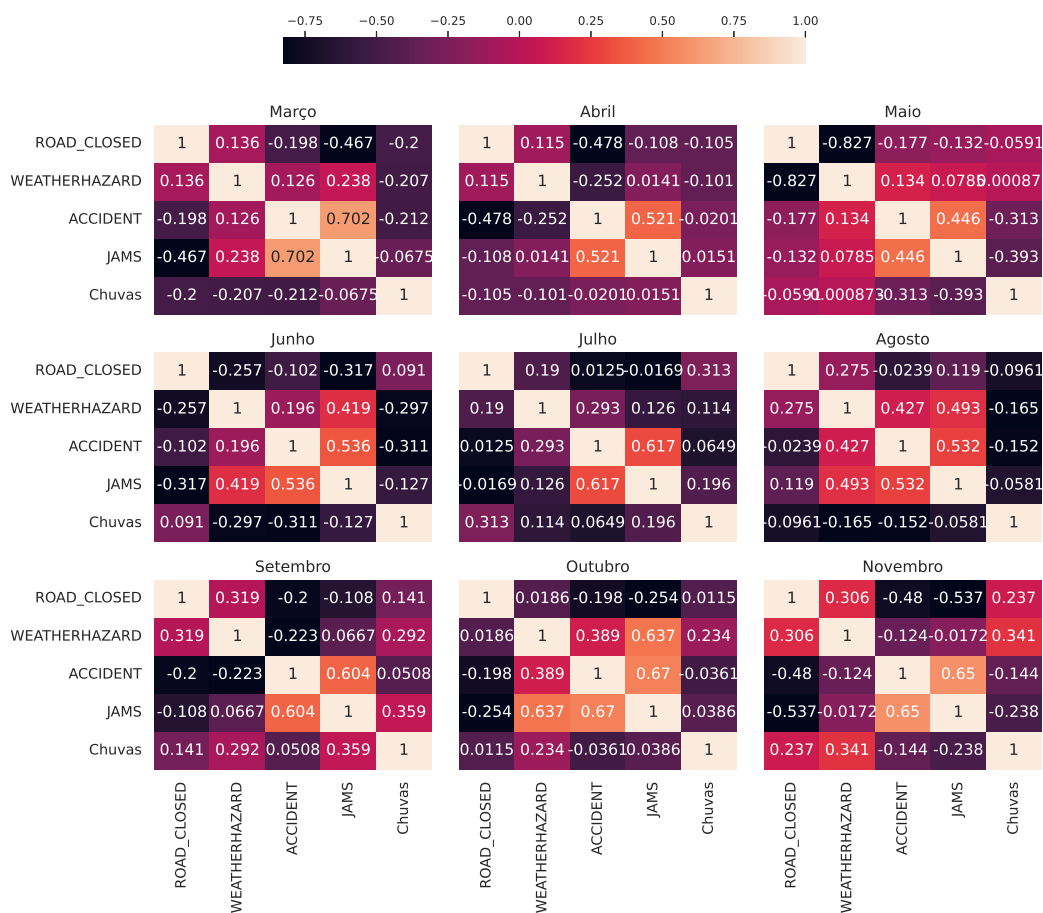


Figura 8. Correlação entre as chuvas e eventos reportados entre Março e Novembro de 2019

Nesse sentido, a figura evidencia que o evento ACCIDENT possui correlações positivas com o evento JAMS durante todos os meses, com uma correlação maior em Março (0,702), enquanto nos outros meses as correlações entre esses eventos variam de $\approx 0,4$ à $\approx 0,6$. Como já mencionado, possivelmente os acidentes são reportados antes e, naturalmente, provocam a ocorrência dos JAMS. É interessante notar, além disso, que há correlação positiva entre os eventos do tipo WEATHERHAZARD e JAMS nos meses de Junho (0,419), Agosto (0,493) e Outubro (0,637). Entretanto, nos meses de Junho e Agosto não há correlação positiva entre as chuvas e o evento WEATHERHAZARD, sugerindo que algum outro tipo de evento, não necessariamente relacionado ao clima, tenha causado o congestionamento do trânsito. Por fim, nota-se que em Agosto há uma correlação positiva entre os eventos ACCIDENT e WEATHERHAZARD (0,427). Assim como no caso anterior, infere-se que as chuvas não causaram os acidentes, pois a correlação entre as chuvas e WEATHERHAZARD nesse mês é negativa.

A partir dos resultados, nota-se que a influência das chuvas nos eventos reportados pelos usuários fica mais evidente no mês de Novembro. Embora outros fatores não considerados aqui possam influenciar os registros de eventos do tipo WEATHERHAZARD, a correlação positiva entre chuvas e o evento nesse mês indica que o clima teve um papel importante na motivação dos registros. Essa hipótese é fortalecida também pelo fato das chuvas terem correlação positiva com o evento ROAD_CLOSED, que possui correlação positiva com WEATHERHAZARD e, provavelmente, foi causado por ele. A mesma situação pode ser vista, com menos intensidade, em Setembro e Outubro.

5. Considerações Finais

Neste trabalho, analisamos a interação dos usuários com um aplicativo de navegação baseado em *crowdsensing* na cidade de Vitória, no Espírito Santo e investigamos como a ocorrência de chuvas na cidade influencia as informações fornecidas pelos usuários do aplicativo. Para isso, utilizamos uma base de dados obtida de uma parceria entre a prefeitura de Vitória e um aplicativo de navegação e dados oficiais do Instituto Nacional de Meteorologia (INMET).

A partir das bases de dados e de dois algoritmos propostos, estudamos padrões em tipos de eventos que podem ter relação com condições climáticas: (i) tempos de duração dos eventos reportados pelos usuários, (ii) o número de eventos reportados nos meses de observação dos dados. Essas análises são importantes para entendermos características dos dados, por exemplo, observamos que, no geral, a maioria dos eventos não apresentam um padrão de sazonalidade durante os meses, podendo estar sendo influenciados por fatores externos, como climáticos. Com isso, medimos a correlação entre as chuvas e os eventos reportados. Observamos que a correlação geral entre as chuvas e os eventos é positiva, com um valor relativamente baixo (0,25), e mais evidente quando considera-se meses individuais.

Como trabalhos futuros, pretende-se aprofundar o estudo considerando cenários mais específicos como, por exemplo, a duração dos eventos em meses e ruas individuais. Além disso, como ficou evidente que as chuvas possuem uma correlação fraca com os dados utilizados, cabe estudar a influência que outros dados externos tenham nos eventos reportados pelos usuários do aplicativo. Por fim, planejamos investigar, também, se regiões (periferia, centros e litoral) da cidade influenciam na forma como os dados são reportados pelos usuários.

Agradecimentos

O presente trabalho foi realizado com apoio do Programa Institucional de Iniciação Científica da UFES, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001, do CNPq e da Fundação de Amparo à Pesquisa do Espírito Santo (FAPES). Adicionalmente, este trabalho foi viabilizado por meio do termo de cooperação técnica 004/2018, entre a Secretaria Municipal de Segurança Pública de Vitória-Espírito Santo e a UFES. Os autores agradecem o esforço da Secretaria pela disponibilização dos dados em tempo real. O trabalho foi também parcialmente apoiado pelo projeto FAPESP-GoodWeb (Processo 2018/23011-1) e pelo CNPq (Processo 310998/2020-4).

Referências

- dos Santos, S. R., Davis Jr, C. A., and Smarzarro, R. (2017). Analyzing traffic accidents based on the integration of official and crowdsourced data. *Journal of Information and Data Management*, 8(1):67–67.
- Gulia, S., Prasad, P., Goyal, S., and Kumar, R. (2020). Sensor-based wireless air quality monitoring network (swaqmn)-a smart tool for urban air quality management. *Atmospheric Pollution Research*, 11(9):1588–1597.
- Ighalo, J. O. and Adeniyi, A. G. (2020). A comprehensive review of water quality monitoring and assessment in nigeria. *Chemosphere*, 260:127569.
- Ismagilova, E., Hughes, L., Dwivedi, Y. K., and Raman, K. R. (2019). Smart cities: Advances in research—an information systems perspective. *International Journal of Information Management*, 47:88–100.
- Joo, H., Ahmed, S. H., and Lim, Y. (2020). Traffic signal control for smart cities using reinforcement learning. *Computer Communications*, 154:324–330.
- Lenkei, Z. (2018). Crowdsourced traffic information in traffic management: Evaluation of traffic information from waze.
- Lin, Y. and Li, R. (2020). Real-time traffic accidents post-impact prediction: Based on crowdsourcing data. *Accident Analysis & Prevention*, 145:105696.
- Pule, M., Yahya, A., and Chuma, J. (2017). Wireless sensor networks: A survey on monitoring water quality. *Journal of applied research and technology*, 15(6):562–570.
- Silva, T. H., Vaz de Melo, P. O. S., Viana, A., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2013). Traffic Condition is more than Colored Lines on a Map: Characterization of Waze Alerts. In *Proc. of the International Conference on Social Informatics (SocInfo'13)*, Kyoto, Japan.
- SM, S. N., Yasa, P. R., Narayana, M., Khadirnaikar, S., and Rani, P. (2019). Mobile monitoring of air pollution using low cost sensors to visualize spatio-temporal variation of pollutants at urban hotspots. *Sustainable Cities and Society*, 44:520–535.
- Solmaz, G., Wu, F.-J., Cirillo, F., Kovacs, E., Santana, J. R., Sánchez, L., Sotres, P., and Munoz, L. (2019). Toward understanding crowd mobility in smart cities through the internet of things. *IEEE Communications Magazine*, 57(4):40–46.
- Thomé, M., Prestes, A., Gomes, R., and Mota, V. (2020). Um arcabouço para detecção e alerta de anomalias de mobilidade urbana em tempo real. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 784–797. SBC.