

Métodos para auxílio na tomada de decisão em cidades inteligentes

Ramon Barros de Lima¹, Wilamis Micael de Araujo Aviz¹,
Fábio J Coutinho¹, João Paulo Clarindo²

¹Instituto de Computação (IC)
Universidade Federal de Alagoas (UFAL)
57072-900 – Maceió – AL – Brazil

²Instituto de Ciências Matemáticas e Computação (ICMC)
Universidade de São Paulo (USP)
13566-590 – São Carlos – SP – Brazil

{rbl,wmaa,fabio}@ic.ufal.br, jpcsantos@usp.br

Abstract. *In recent years, with the population increase in urban areas, smart cities have emerged as an alternative to assist managers in decision-making, through the use of an IoT device network, which generate data related to urban infrastructure. However, the number of IoT devices is constantly growing, resulting in a huge volume of conventional and spatial data. Therefore, we discuss methods to assist decision-making in smart cities context, through spatial data pre-processing techniques, parallel and distributed processing of IoT devices data. We developed a case study based on data from 500,000 taxi rides that took place in Porto, Portugal, where we present results that can help managers to improve the city's transport services.*

Resumo. *Nos últimos anos, com o crescente aumento da população em áreas urbanas, cidades inteligentes surgem como uma alternativa para auxiliar gestores na tomada de decisão, através da utilização de uma rede de dispositivos IoT, que geram dados relativos à infraestrutura urbana. Entretanto, o número de dispositivos IoT cresce constantemente, gerando um imenso volume de dados convencionais e espaciais. Diante disso, este trabalho discute métodos para auxiliar na tomada de decisão no contexto de cidades inteligentes, a partir do uso de técnicas de pré-processamento dos dados brutos gerados por dispositivos IoT, armazenamento e processamento paralelo e distribuído de dados espaciais. Foi desenvolvido um estudo de caso com dados de cerca de 500 mil corridas de táxi ocorridas em Porto, Portugal, que apresenta resultados que podem auxiliar gestores na melhoria do serviço de transportes da cidade.*

1. Introdução

Nos últimos anos, a população disposta em áreas urbanas tem crescido constantemente. Segundo a Organização das Nações Unidas (ONU) (2018), 68% da população irá viver em áreas urbanas em 2050. Diante disso, lidar com problemas urbanos em cidades com grande densidade populacional tem sido um desafio importante para órgãos governamentais. Segundo Tibari e Mazri (2019), cidades inteligentes (ou *smart cities*) representam “um modelo de cidade onde sua infraestrutura está ligada ao desenvolvimento urbano

economicamente sustentável”. Neste contexto, cidades inteligentes relacionam-se com dispositivos e sensores capazes de coletar dados do ambiente [Yeh 2017] para apoio à tomada de decisão, auxiliando na melhoria da infraestrutura e da qualidade de vida da população.

Dentre os componentes que integram o ecossistema de uma cidade inteligente, encontram-se os dispositivos de Internet das Coisas (IoT) ou *Internet of Things*, que segundo Patel and Patel (2016), consistem em objetos interconectados que geram dados regularmente analisados de modo a prover informações para gerência e tomada de decisões. Os dados gerados por dispositivos IoT podem abordar diferentes questões tais como transporte público, consumo de energia, poluição, entre outras. Além disso, são provenientes de uma grande quantidade de locais e fontes, gerando um enorme volume de dados heterogêneos sejam convencionais ou espaciais [Atzori et al. 2017, Yeh 2017, Eldrandaly et al. 2019].

No entanto, as soluções tradicionais não se mostram eficazes para o armazenamento, processamento e gerência desses dados. Neste contexto, serviços de *big data* oferecem velocidade de processamento de consultas e conseguem lidar com variedade de dados estruturados e não-estruturados [Ishwarappa and Anuradha 2015]. Existem diversas tecnologias disponíveis para o auxílio na tomada de decisão em *big data*, como sistemas de processamento e armazenamento paralelo e distribuído. Algumas dessas tecnologias não oferecem suporte para a manipulação de dados espaciais, necessitando de Sistemas Analíticos Espaciais (SAEs) para possibilitar a execução de consultas analíticas espaciais [Castro et al. 2020].

Neste trabalho são discutidos métodos para lidar com dados espaciais gerados por dispositivos IoT no contexto de cidades inteligentes. Os métodos foram criados seguindo o fluxo de extração, transformação, carga, análise e visualização. Para avaliação da proposta foi realizado um estudo de caso que analisou consultas espaciais utilizando o *framework* Apache Spark¹ e Apache Sedona² para processamento paralelo e distribuído, e Apache Hadoop HDFS³ para armazenamento. Os resultados foram dispostos em mapas e gráficos e revelaram padrões interessantes para possível aplicação em projetos de mobilidade urbana.

O artigo encontra-se organizado da seguinte forma. Na Seção 2 são discutidos os trabalhos relacionados. Na Seção 3 são apresentados os métodos para análise e visualização de dados espaciais no contexto de cidades inteligentes. Na Seção 4 é apresentado um estudo de caso que implementa os métodos da solução proposta na seção anterior. Por fim, na Seção 5, são apresentadas as considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

Nesta seção, são discutidos trabalhos existentes na literatura que manipulam dados espaciais gerados por dispositivos IoT no contexto de cidades inteligentes.

Huneiti and Adwan (2019) utilizaram uma rede neural do tipo *Self Organizing Map* (SOM) para detectar Regiões de Interesse (ROI) da cidade de Porto a partir do agrupamento (*clustering*) de coordenadas GPS de embarques e desembarques de corridas de

¹<https://spark.apache.org/>

²<https://sedona.apache.org/>

³<https://hadoop.apache.org/>

táxis extraídas de dados de trajetórias dos veículos publicadas em um *dataset* público. Já Liu et al. (2020) investigaram a correlação entre fatores socioeconômicos e a demanda de táxis em diferentes regiões da cidade de Qingdao na China utilizando o modelo de regressão geográfica ponderada (GWR). Dentre as conclusões deste último, verificou-se que em áreas residenciais é maior o número de solicitações de táxis em regiões com imóveis de valor elevado.

O trabalho de Nguyen et al. (2018) descreve um sistema que dispõe de um conjunto de métricas acerca da estimativa de chegada, atraso, confiabilidade e incidência de engarrafamento, visando a melhoria dos serviços de transporte urbano. Como estudo de caso, os autores utilizaram dados espaciais referentes a cerca de 38 mil viagens de ônibus pertencentes ao Metrô de Los Angeles, disponíveis em um *dataset* público⁴, gerados a partir de sensores contidos na frota. Por fim, Amaral et al. (2016) utilizaram dados de trajetórias obtidos de dispositivos IoT implantados na frota de ônibus da cidade do Rio de Janeiro com o intuito de criar a ferramenta BusesInRio para disponibilizar dados em tempo real em mapas, indicando posição geográfica, linhas de ônibus e tempo de viagem. Os dados, que são abertos, foram obtidos a partir do portal *datario*⁵.

Em todos os trabalhos anteriormente discutidos foram analisados dados espaciais relativos à trajetória de veículos (corridas de táxi e linhas de ônibus). Entretanto, os autores não utilizaram processamento paralelo e distribuído para a execução das consultas espaciais. Diferentemente, neste trabalho propõe-se o uso de *frameworks* para o processamento paralelo e distribuído de dados espaciais no contexto de cidades inteligentes.

3. Métodos para análise e visualização de dados espaciais gerados por dispositivos IoT

Para viabilizar a análise de dados espaciais gerados por dispositivos IoT no contexto de cidades inteligentes, inicialmente, deve-se definir como os dados são extraídos, armazenados e carregados em um ambiente de armazenamento paralelo e distribuído; em seguida, como serão analisados e visualizados [Santos et al. 2020]. A Figura 1 ilustra um fluxo de dados para a extração, transformação, análise e visualização de dados gerados por dispositivos IoT no contexto de cidades inteligentes, dividido nos seguintes passos: (i) extração e transformação dos dados brutos; (ii) carga dos dados transformados em um sistema de arquivos paralelo e distribuído; (iii) análise dos dados em um ambiente de processamento paralelo e distribuído, utilizando SAEs para a execução de consultas espaciais; e (iv) visualização dos resultados. As seções seguintes descrevem cada etapa do fluxo de dados ilustrado na Figura 1.

3.1. Método de Extração e Transformação

Nesta etapa, é realizada a extração e transformação de dados gerados por dispositivos IoT mediante a implementação de programas que utilizam bibliotecas de suporte à manipulação e transformação de dados espaciais. Um exemplo de transformação consiste em converter dados de latitude e longitude referentes a trajetórias em objetos *Well-Known Text* (WKT) do tipo *linestring*. Algumas operações básicas de transformação são voltadas para a limpeza dos dados tais como a eliminação de dados inconsistentes, nulos e repetidos.

⁴<https://developer.metro.net/gtfs-schedule-data/>

⁵<https://www.data.rio/>

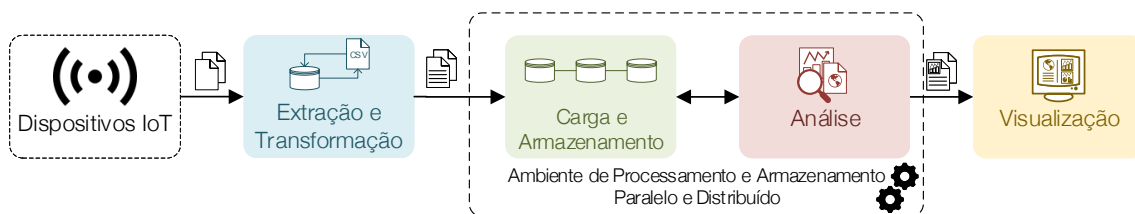


Figura 1. Fluxo de dados para a extração, transformação, carga e armazenamento, análise e visualização de dados gerados por dispositivos IoT no contexto de cidades inteligentes.

3.2. Método de Carga e Armazenamento

Após a extração e transformação dos dados, os arquivos com os dados transformados são dispostos de forma estruturada para carregamento em um ambiente de armazenamento paralelo e distribuído, mantido em um *cluster* local ou em serviços de nuvem, como *Storage-as-a-service* (STaaS). Esses ambientes de armazenamento devem prever replicação, alta disponibilidade e suporte ao armazenamento de dados estruturados de diferentes formas.

3.3. Método de Análise

Com os dados transformados e carregados no ambiente de armazenamento, a etapa de análise inicia-se pela escolha do *framework* para o processamento paralelo e distribuído de consultas espaciais. Consultas com dados convencionais incluem agrupamento e cálculos estatísticos. Considerando a análise de dados espaciais, exemplos de consultas incluem: (i) *End Point Query*, onde dado um objeto espacial, a consulta retorna o último ponto contido em uma linha; (ii) *Touch Query*, onde dado um par de objetos espaciais, a consulta retorna verdadeiro caso esses objetos tenham ao menos um ponto em comum; (iii) *Buffer Query*, onde dado um objeto espacial o , a consulta retorna um polígono de largura (ou raio) n , que circunda o . As operações de análise podem ser executadas utilizando linguagens de programação ou linguagens de consultas, conforme o suporte oferecido pelo *framework* de processamento paralelo e distribuído.

3.4. Método de Visualização

Após os dados serem extraídos, transformados, armazenados e analisados, os resultados da análise podem ser exibidos através de ferramentas de visualização, tais como *dashboards*, relatórios, e mapas, utilizando Sistemas de Informações Geográficas (SIGs). A disposição dos resultados da análise de dados espaciais em ferramentas de visualização permite aos gestores de cidades inteligentes identificar padrões regionais, auxiliando-os no processo de tomada de decisão.

4. Estudo de caso

Seguindo os métodos descritos na Seção 3, foi desenvolvido um estudo de caso que utiliza um *dataset*⁶ contendo dados relacionados a corridas de 442 táxis da cidade de Porto, em Portugal. Esses táxis operam através de uma central, utilizando terminais móveis de dados instalados nos veículos, que geram dados relativos à identificação da corrida, forma utilizada para solicitar o serviço de viagem, e uma lista de coordenadas geradas a cada 15 segundos. Foram utilizadas 500.000 amostras do conjunto de dados coletados entre 01/07/2013 e 30/06/2014.

⁶<https://www.kaggle.com/crailltap/taxi-trajectory/metadata>

Para a limpeza e transformação dos dados convencionais e espaciais, foi utilizada a linguagem de programação Python, com a biblioteca GeoPandas⁷. Para o processamento e armazenamento paralelo distribuído, foi utilizado o Apache Spark e Apache Hadoop HDFS. Como o Spark não provê o suporte nativo a consultas espaciais, foi escolhido o SAE Apache Sedona (antigo GeoSpark), seguindo as diretrizes descritas por Castro et al. (2020). As consultas foram realizadas utilizando SparkSQL. Para a visualização dos dados, foram utilizadas a biblioteca para Python Matplotlib⁸ para a criação de gráficos, e o SIG QGIS⁹ para a exibição dos dados espaciais. As seções seguintes descrevem algumas consultas de interesse do planejamento da mobilidade urbana em cidades inteligentes cujos resultados estão dispostos em gráficos ou mapas.

4.1. Densidade de embarques conforme o dia da semana

A Figura 2 ilustra a incidência de embarques considerando todas as corridas realizadas nas segundas-feiras identificando as regiões da cidade de Porto com maior demanda por táxi nesse dia da semana. A área central da cidade apresentou um maior quantitativo de embarques.

```
SELECT ST_StartPoint(t.geometry) AS geometry, t.week_day  
FROM travels t WHERE t.week_day = 0
```

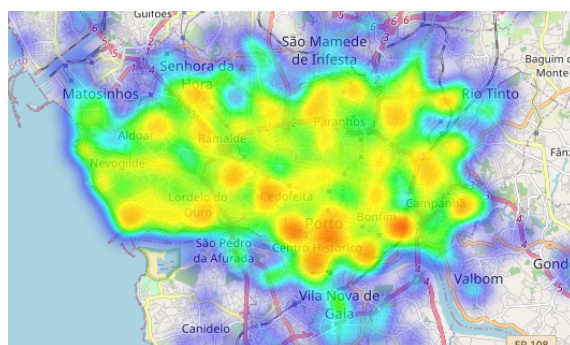


Figura 2. Gráfico de calor apresentando a densidade de embarques nas segundas

4.2. Identificando origem ou destino de corridas

Identificar áreas com maior número de embarques e desembarques de corridas de táxi pode ser útil para conhecer as regiões da cidade de maior interesse. Tal informação é relevante em projetos de mobilidade urbana. A consulta a seguir retorna o último ponto pertencente a uma linestring que representam corridas de táxi.

```
SELECT ST_EndPoint(t.geometry) as geometry, t.TRIP_ID  
FROM travels t  
LIMIT 1000
```

A Figura 3a apresenta os pontos de desembarque retornados pela consulta anterior plotados no mapa da cidade de Porto considerando uma amostra de 1000 corridas de táxi. Percebe-se uma maior incidência de desembarques na área central da cidade.

⁷<https://www.geopandas.org>

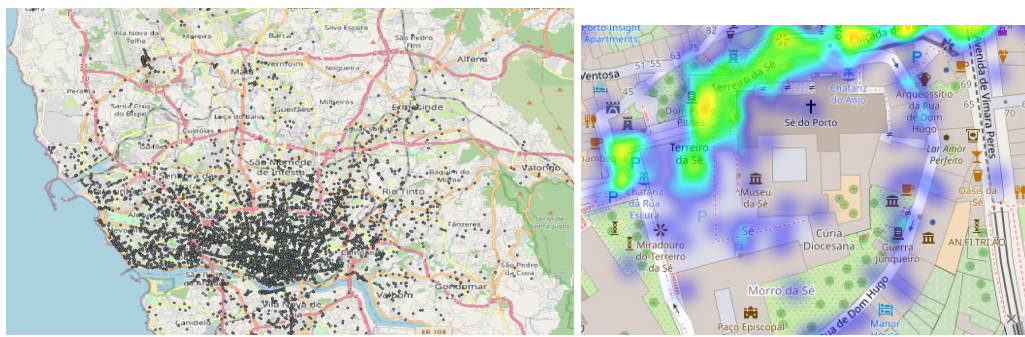
⁸<https://matplotlib.org/>

⁹<https://qgis.org>

4.3. Incidência de desembarques próximos a pontos de interesse

Para obter os pontos de interesse da cidade de Porto foi utilizada uma base auxiliar extraída do OpenStreetMap¹⁰. A consulta apresentada a seguir relaciona as corridas de táxi com os pontos de interesse recuperados do OpenStreetMap, retornando onde ocorrem desembarques que tenham ao menos um ponto em comum com a região formada a partir da aplicação de um *buffer* de 5 metros de diâmetro ao Ponto de Interesse (POI). A Figura 3b ilustra o fluxo de desembarque em volta do POI “Sé do Porto” a partir de um mapa de calor.

```
SELECT ST_EndPoint(t.geometry) AS geometry, p.name, t.TRIP_ID
FROM travels t, points_of_interest p
WHERE ST_Touches(ST_EndPoint(t.geometry), ST_Buffer(p.geometry, 5))
```



(a) Pontos de desembarque de corridas de táxi na cidade de Porto. (b) Mapa de calor representando a incidência de desembarques próximos ao ponto de interesse “Sé do Porto”.

Figura 3. Resultados das consultas relacionadas à incidência de corridas de táxi na cidade de Porto, Portugal.

5. Considerações finais

Este trabalho discutiu possíveis métodos para o processamento, análise e visualização de dados espaciais gerados por dispositivos IoT no contexto de cidades inteligentes. Na revisão da literatura, encontram-se trabalhos que manipulam dados espaciais, entretanto, nenhum deles faz uso de SAEs para o processamento paralelo e distribuído de consultas espaciais.

Foi realizado um estudo de caso para implementar os métodos de análise e visualização a partir da criação de consultas espaciais com o intuito de auxiliar à tomada de decisão por parte de gestores de cidades inteligentes. Os resultados identificaram padrões interessantes para possível aplicação em projetos de mobilidade urbana.

Como trabalhos futuros, pretende-se considerar novos estudos de caso incluindo o processamento e análise de dados gerados em tempo real. Além disso, serão aplicadas novas métricas para a avaliação do desempenho da execução dos métodos implementados.

¹⁰<https://www.openstreetmap.org/>

Referências

- Amaral, B. G. D., Nasser, R., Casanova, M. A., and Lopes, H. (2016). Busesinrio: Buses as mobile traffic sensors: Managing the bus gps data in the city of rio de janeiro. *Proceedings - IEEE International Conference on Mobile Data Management*, pages 369–372.
- Atzori, L., Iera, A., and Morabito, G. (2017). Understanding the internet of things: definition, potentials, and societal role of a fast evolving paradigm. *Ad Hoc Networks*, 56:122–140.
- Castro, J. P., Carniel, A., and Ciferri, C. (2020). Analyzing spatial analytics systems based on Hadoop and Spark: A user perspective. *Software: Practice and Experience*, 50(12):2121–2144.
- Eldrandaly, K. A., Abdel-Basset, M., and Shawky, L. A. (2019). Internet of Spatial Things: A New Reference Model With Insight Analysis. *IEEE Access*, 7:19653–19669.
- Huneiti, A. M. and Adwan, O. Y. (2019). Extracting interesting regions and trips from taxi trajectory data. *Modern Applied Science*, 13:258.
- Ishwarappa and Anuradha, J. (2015). A brief introduction on big data 5vs characteristics and hadoop technology. *Procedia Computer Science*, 48:319–324. International Conference on Computer, Communication and Convergence (ICCC 2015).
- Liu, X., Sun, L., Sun, Q., and Gao, G. (2020). Spatial variation of taxi demand using gps trajectories and poi data. *Journal of Advanced Transportation*, 2020.
- Nguyen, K., Yang, J., Lin, Y., Lin, J., Chiang, Y.-Y., and Shahabi, C. (2018). Los angeles metro bus data analysis using gps trajectory and schedule data (demo paper). In *SIGSPATIAL 18: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 560–563.
- ONU (2018). 68% of the world population projected to live in urban areas by 2050, says un.
- Patel, K. K. and Patel, S. M. (2016). Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges. *IJSR*, 6(5):6122–6132.
- Santos, J. P. C., Castro, J. P. d. C., and Ciferri, C. D. d. A. (2020). SOLAP Query Processing over IoT Networks in Smart Cities: A Novel Architecture. In *Anais do XXI GeoInfo - Simpósio Brasileiro de Geoinformática*, pages 118–129, São José dos Campos, Brazil. INPE.
- Tibari, S. and Mazri, T. (2019). Security in a smart city: Challenges and solutions. *ACM International Conference Proceeding Series*.
- Yeh, H. (2017). The effects of successful ict-based smart city services: From citizens' perspectives. *Government Information Quarterly*, 34:556–565.