SILVIA GOLDMAN BER KAPEL

**Análise de dados agregados aplicada à mobilidade urbana – Um framework
analítico para a geração de conhecimento**

São Paulo
(2020)

SILVIA GOLDMAN BER KAPEL

**Análise de dados agregados aplicada à mobilidade urbana – Um Framework
Analítico para geração de conhecimento**

Versão Corrigida

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Ciências

Área de Concentração: Engenharia de
Computação

Orientador: Prof. Dr. Jorge Rady de
Almeida Junior

São Paulo
(2020)

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 27 de Novembro de 2020.

Assinatura do autor: S. G. B. Kapel

Assinatura do orientador: Prof. Dr. Alexandre F. de Almeida

CATALOGAÇÃO-NA-PUBLICAÇÃO

Kapel, Silvia Goldman Ber

Análise de dados agregados aplicada à mobilidade urbana – Um framework analítico para a geração de conhecimento / S. G. B. Kapel -- São Paulo, 2020.

78 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Análise de Dados 3.Mobilidade Urbana 4.Dinâmica Humana
5.Framework I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

DEDICATÓRIA

.

Ao meu marido e filhos que sempre me apoiaram e que, compreensivamente, souberam respeitar minha dedicação aos estudos.

À minha querida mãe, que sempre esteve ao meu lado, ajudando a me tornar a pessoa que sou hoje.

Ao meu primo Alfredo e tio David, que tanto me incentivaram nessa trajetória.

AGRADECIMENTOS

Meus agradecimentos ao Prof. Dr. Jorge Rady de Almeida Junior, pela orientação e apoio na realização desse trabalho. Agradeço também em especial por sua confiança, dedicação e incentivo.

Agradeço aos demais professores da Escola Politécnica que por suas disciplinas ou encontros pontuais me proporcionaram as bases acadêmicas para chegar até aqui.

Meus agradecimentos também aos colegas de disciplinas que compartilharam comigo experiências que foram essenciais para a realização desse trabalho.

Gostaria de expressar meu agradecimento à minha amiga Sara, que me incentivou a iniciar essa jornada.

RESUMO

KAPEL, S.G.B. Análise de dados agregados aplicada à mobilidade urbana – Um Framework Analítico para geração de conhecimento. 2020. Dissertação (Mestrado em Ciências) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2020.

As cidades têm enfrentado grandes desafios com o crescimento da população urbana e o aumento da demanda de recursos de infraestrutura, exigindo que busquem por soluções que as tornem mais inteligentes. Um importante alavancador dessas soluções é a proliferação de conteúdo digital das diversas fontes de dados, que gera uma oportunidade única de análise e extração de conhecimento no contexto de Mobilidade Urbana, em Cidades Inteligentes. Neste cenário, a evolução da Internet das Coisas, com seus sensores e *smartphones*, tem contribuição de destaque na geração de dados para melhor compreensão da mobilidade urbana. Esse conteúdo digital é muito rico, trazendo informações em larga escala e com alto nível de detalhes da dinâmica humana urbana. Essas informações oferecem inúmeras possibilidades de aplicação, onde a análise dos dados gera subsídios para maior conhecimento sobre o comportamento humano nas cidades. Por outro lado, a análise desses fluxos de dados apresenta diversos desafios para a extração do conhecimento desejado no tempo necessário. Muitos são os métodos e técnicas utilizados para extrair valor desses dados em aplicações de mobilidade urbana, com questões a serem endereçadas em cada uma das etapas que compõem o processo de análise, não havendo hoje a compilação de um conjunto de recomendações a serem seguidas ou a serem evitadas para atingir-se os resultados esperados no tempo desejado. Esta pesquisa tem por objetivo reunir, em um framework analítico, um conjunto de melhores práticas para análise de dados provenientes de telefonia móvel, aplicada à mobilidade urbana. Para tal é traçado um panorama sobre o triângulo formado por mobilidade urbana, dinâmica humana e dados de telefonia móvel seguido de estudo de caso contido nas referências, avaliando-se os métodos e técnicas utilizados, dificuldades encontradas, pontos fortes e resultados obtidos. Finalmente propõe-se o desenvolvimento de um framework que contemple essas melhores práticas.

Palavras-Chave: Análise de Dados. Cidades Inteligentes. Mobilidade Urbana. Dinâmica Humana. Telefonia Móvel. Framework.

ABSTRACT

KAPEL, S.G.B. Aggregated data analytics for urban mobility – An Analytics framework for knowledge generation. 2020. Dissertação (Mestrado em Ciências) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2020.

Nowadays cities have faced great challenges with urban population growth and increased demand for infrastructure resources, requiring them to search for solutions to make them more intelligent. An important lever of these solutions is proliferation of digital content from several data sources, which generates a unique opportunity for analysis and knowledge extraction in the context of Urban Mobility, in Smart Cities. In this scenario, the evolution of the Internet of Things, with its meters, sensors and smartphones, provides an outstanding contribution to the generation of data of great value for a better understanding of urban mobility. Rich digital content from the most diverse sources brings information on a large scale and with a high level of details on urban dynamics and human activity. This information offers countless usage possibilities, where the analysis of the data, both in real time, as well as historical, generates subsidies for a more intelligent transport and enables greater knowledge about human behavior in cities. On the other hand, the analysis of these data flows presents several challenges for the timely extraction of the desired knowledge. There are many methods and techniques used to extract value from these data in Urban Mobility applications, with questions to be addressed in each of the steps that make up the analysis process, and today there is no compilation of a set of recommendations to be followed or avoided to achieve the expected results in the desired time. This research aims to gather, in an analytical framework, a set of best practices for analyzing data from mobile phone, applied to urban mobility. To this end, a panorama is drawn on the triangle formed by Urban Mobility, Human Dynamics and Mobile Phone Data followed by a case study contained in the references, evaluating the methods and techniques used, difficulties encountered, strengths and results obtained. Finally, it is proposed a framework that includes best practices.

Keywords: Data Analytics. Smart Cities. Urban Mobility. Human Dynamics. Mobile Phone. Framework.

LISTA DE ILUSTRAÇÕES

Figura 1 - Os 5 V's de Big Data.....	18
Figura 2 - Cidade inteligentes: Um sistema de sistemas.....	25
Figura 3 - Etapas do macro-processo de cidades inteligentes.....	26
Figura 4 - Ciclo contínuo de melhorias.....	27
Figura 5 - Ciclo completo de análise de dados.....	28
Figura 6 - Etapas do desenvolvimento da pesquisa.....	34
Figura 7 - Modelo de Observação.....	44
Figura 8 - Framework analítico para mobilidade urbana	64
Figura 9 - Crescimento do número de assinantes de telefonia móvel e taxa de penetração	69
Figura 10 - Evolução dos assinantes de telefonia móvel nos países em desenvolvimento	70

LISTA DE QUADROS E TABELAS

Quadro 1- Desafios de análise de dados em mobilidade urbana	30
Quadro 2 - Bibliotecas digitais	38
Quadro 3 - Lista de referências selecionadas	40
Quadro 4 – Sumário das áreas de aplicação em mobilidade urbana	45
Quadro 5 - Sumário da Tipificação dos dados da rede de telefonia móvel	54
Quadro 6 - Sumário Pré-Processamento	55
Quadro 7 - Sumário Análise	57

SUMÁRIO

1 INTRODUÇÃO	13
1.1 MOTIVAÇÃO	13
1.2 OBJETIVO DA PESQUISA	14
1.3 MÉTODO	15
1.4 ORGANIZAÇÃO DO DOCUMENTO	15
2 REVISÃO DA LITERATURA	17
2.1 A DEFINIÇÃO DE BIG DATA	18
2.2 DATA ANALYTICS	19
2.3 DADO ESTRUTURADO E DADO NÃO-ESTRUTURADO	20
2.3.1 A Importância e os desafios da análise do dado não-estruturado	20
2.4 INTERNET DAS COISAS	22
2.5 CIDADES INTELIGENTES E MOBILIDADE URBANA	23
2.6 ANÁLISE DE DADOS E MOBILIDADE URBANA: AS OPORTUNIDADES	24
2.7 ANÁLISE DE DADOS E MOBILIDADE URBANA: OS DESAFIOS	27
2.7.1 O processo de análise de dados	27
2.7.2 Os desafios para obtenção dos resultados desejados	29
2.8 TELEFONIA MÓVEL COMO FONTE DE DADOS	31
3 ANÁLISE DE DADOS DE MOBILIDADE URBANA – MÉTODO DE TRABALHO	33
3.1 VISÃO GERAL	33
3.2 ETAPAS DO DESENVOLVIMENTO DA PESQUISA	34
3.2.1 Etapa 1 – Seleção das referências	34
3.2.2 Etapa 2 – Modelo de observação	35
3.2.3 Etapa 3 – Observação das referências	35
3.2.4 Etapa 4 – Avaliação dos resultados	35
3.2.5 Etapa 5 – Recomendações	36
4 ANÁLISE DE DADOS DE MOBILIDADE URBANA – DEFINIÇÃO DE UM FRAMEWORK ANALÍTICO	37
4.1 SELEÇÃO DAS REFERÊNCIAS	37
4.2 MODELO DE OBSERVAÇÃO	43
4.2.1 Áreas de aplicação em mobilidade urbana	44
4.2.2 Tipificação da contribuição dos dados da rede de telefonia móvel	50
4.2.3 Técnicas de pré-processamento e análise	54
4.3 AVALIAÇÃO DOS RESULTADOS	59
4.4. DEFINIÇÃO DE UM FRAMEWORK ANALÍTICO	62
5 DISCUSSÃO	69
6 CONCLUSÃO	72

1 INTRODUÇÃO

A evolução da humanidade tem caminhado cada vez mais na direção da vida urbana. Atualmente, mais de metade da população vive em ambientes urbanos, demandando grandes esforços para que sejam criadas condições de vida viáveis. Ainda, o avanço da tecnologia de comunicação pessoal móvel tem contribuído para o grande crescimento do seu número de usuários, que mais que dobrou nos últimos dez anos, representando uma grande parcela da população mundial, e trazendo dados de abrangência sem igual para conhecimento da dinâmica humana nos centros urbanos, tornando-se uma importante área de estudo (Vilajosana et al., 2013) (Nations, of Economic, Affairs, & Division, 2018) (“International Telecommunication Union (ITU) Statistics,” n.d.).

Este capítulo apresenta a motivação para a pesquisa realizada, seu objetivo, o sumário do método utilizado e a organização do documento.

1.1 MOTIVAÇÃO

Em função do crescimento da população nos centros urbanos, é urgente a busca por novas soluções, e os avanços tecnológicos podem ser de grande ajuda para endereçar os problemas enfrentados pelas cidades.

Neste cenário, é crescente a procura por soluções de cidades inteligentes como resposta aos desafios atuais na vida urbana, e a área de mobilidade urbana é um importante foco de atuação. Um dos principais alavancadores de soluções de cidades inteligentes, em particular em mobilidade urbana é a análise de dados. A proliferação de conteúdo digital das diversas fontes de dados como sensores, mídias sociais, smartphones, páginas web, e-mails, áudio, imagens e outros, denominados como Big Data, gera uma oportunidade única de extração de conhecimento.

De acordo com (Fisher, DeLine, Czerwinski, & Drucker, 2012), a análise de dados é considerada a principal tendência de tecnologia da informação do século 21, abrindo grandes oportunidades para lidar-se com problemas sociais, econômicos e de negócio. A análise de dados é vista como uma fonte inexplorada de melhorias de resultados que, se usado de forma apropriada, será a chave para elevar as cidades a um novo patamar de gestão de seus recursos.

Um dos mais importantes e maior gerador de dados no contexto de cidades inteligentes é Internet das coisas ou Internet of Things (IoT) em inglês. O poder dos dados fornecidos por todos os recursos que estão sendo conectados à Internet traz uma nova concepção das cidades e de seu relacionamento com os cidadãos (Jara, Genoud, & Bocchi, 2014).

Em IoT, os dispositivos pessoais de telefonia móvel, em particular, ocupam um lugar de extrema relevância, trazendo dados sobre a população de abrangência sem igual, sendo de grande valor para a área de mobilidade urbana (Naboulsi, Fiore, Ribot, & Stanica, 2016) (Calabrese, Ferrari, & Blondel, 2014). Estes dados possuem duas grandes principais fontes geradoras, a primeira correspondente aos registros relacionados ao uso da rede de telefonia móvel, e a segunda aos dados gerados por aplicativos instalados nos smartphones.

O tratamento efetivo dos dados gerados por esses dispositivos pessoais de telefonia móvel requer, além do uso de tecnologias avançadas e recursos de processamento rápido, a implementação de novas técnicas e modelos de análise de dados, que de fato agregarão a capacidade de obter novos conhecimentos para atingir-se os resultados almejados em prol das cidades e seus cidadãos.

As cidades que desejam se beneficiar desta nova tendência deverão, ativamente, perseguir a arte e a ciência de capturar, analisar e criar valor a partir dos dados.

Pode parecer simples, mas os desafios para esta manipulação são diversos, desde a definição do escopo de dados, forma de captura e tratamento, as diferentes técnicas para análise, os diferentes passos a serem seguidos, a existência dos perfis adequados na organização para orientar e realizar as análises e inúmeros outros. Não obstante as dificuldades, esta é uma forte necessidade das organizações pois o valor adicionado por estas análises abre portas para um conhecimento hoje não existente (Fan, Han, & Liu, 2014).

1.2 OBJETIVO DA PESQUISA

A pesquisa proposta tem como objetivo avaliar os métodos e técnicas que vêm sendo utilizados para a análise de dados agregados proveniente do uso da rede de telefonia móvel aplicados à área de mobilidade urbana, reunindo um conjunto de melhores práticas que possa ser aplicado às diferentes subáreas dessa mobilidade,

como transportes e dinâmica humana, englobando as várias etapas que compõem o processo de análise de dados, indo da captura dos dados até a obtenção do conhecimento.

Como resultado este estudo apresenta um framework analítico orientado a dados composto por quatro passos, iniciando-se na coleta dos dados, passando pelo pré-processamento para a preparação dos dados para o passo seguinte, a análise propriamente dita, até os resultados. Este framework é acompanhado de um conjunto de recomendações tanto no sentido positivo, isto é, sobre o que se recomenda adotar, como negativo referente às recomendações do que deve ser evitado. Estas recomendações têm como abrangência o ciclo completo de análise de dados, cobertos pelos quatro passos do framework analítico.

As análises de dados contempladas no objetivo desta pesquisa consideram os casos com foco em dados agregados provenientes do uso da rede de telefonia móvel.

1.3 MÉTODO

O método a ser usado para desenvolvimento da pesquisa é baseado no método de Estudo de Caso, exposto por Robert Yin (Yin, 2003), utilizando como fontes de informação artigos de periódicos e conferências, dissertações e teses, consultas a livros e informações disponíveis no meio acadêmico ou públicas.

Para a realização da pesquisa foram selecionadas diversas referências de aplicação de análise de dados em mobilidade urbana, utilizando as fontes anteriormente mencionadas, com foco em fontes de dados proveniente do uso da rede de telefonia móvel. A partir das observações efetuadas nas referências estudadas em relação aos métodos e técnicas adotados, resultados obtidos e dificuldades encontradas, foi compilado o conjunto de recomendações adequado às diferentes possibilidades de análise de dados em mobilidade urbana.

1.4 ESTRUTURA DO TRABALHO

Este primeiro capítulo fornece uma introdução sobre o tema, no segundo capítulo é apresentada uma revisão da literatura envolvendo os conceitos de Big Data, Análise de dados, IoT, cidades inteligentes e mobilidade urbana fornecendo um panorama do triângulo formado por IoT, Big Data e mobilidade urbana, bem como as

oportunidades presentes neste ecossistema, no contexto deste trabalho. No terceiro capítulo são apresentados o método e etapas utilizadas para a realização da pesquisa. O quarto capítulo expõe os resultados obtidos segundo as etapas indicadas no método adotado para a realização da pesquisa, onde é feita a observação das referências selecionadas, e é apresentado o framework analítico orientado a dados acompanhados das recomendações. O capítulo quinto apresenta a discussão dos resultados obtidos com a pesquisa. A conclusão é apresentada no capítulo sexto, seguida pelas referências bibliográficas utilizadas neste trabalho.

2 REVISÃO DA LITERATURA

Os avanços e a convergência das tecnologias da informação e comunicação revolucionaram o modo de vida das pessoas (Naphade, Banavar, Harrison, Paraszczak, & Morris, 2011). O número crescente de assinantes de telefonia móvel torna a tecnologia de comunicação móvel pessoal uma das inovações mais bem-sucedidas dos últimos tempos. Uma grande quantidade de indivíduos depende inteiramente de seus dispositivos móveis, não apenas para o trabalho, mas também para sua vida pessoal. Hoje, os assinantes de telefonia móvel representam uma extensa parcela da população mundial ("International Telecommunication Union (ITU) Statistics," n.d.); seus telefones celulares estão sempre interagindo com a rede de telecomunicações, gerando tráfego e eventos georreferenciados.

Nesse cenário, os rastros digitais gerados pelos usuários de telefones celulares emergiram rapidamente como fonte principal de conhecimento sobre a dinâmica humana (Huang, Cheng, & Weibel, 2019), a um custo mínimo, representando uma importante oportunidade para a extração de conhecimento na área de mobilidade urbana. Compreender a dinâmica humana e sua mobilidade é essencial para muitas aplicações relacionadas à cidade, como planejamento urbano (De Nadai et al., 2016), estudos demográficos (Pappalardo, Pedreschi, Smoreda, & Giannotti, 2015), transporte (Huang et al., 2019), estimativa de fluxos migratórios, gerenciamento de multidões (Celes, Boukerche, & Loureiro, 2019), modelagem epidêmica e previsão de demanda de energia (Selvarajoo, Schlapfer, & Tan, 2018) e outras.

As soluções de mobilidade urbana envolvem dois grandes blocos, o sensoriamento urbano e a análise de dados urbanos. O sensoriamento urbano permite obter dados desses rastros digitais, e a análise possibilita compreender a dinâmica das cidades (Celes et al., 2019). Existem vários métodos e técnicas sobre como coletar, processar e analisar todos esses dados (Calabrese et al., 2014) (Zhao, Tarkoma, Liu, & Vo, 2016) (Naboulsi et al., 2016).

Esta é uma área de grande atividade de pesquisa tanto no meio acadêmico como corporativo.

A seguir é apresentada uma revisão da literatura referente aos principais temas envolvidos.

2.1 A DEFINIÇÃO DE BIG DATA

Segundo Wikipedia (Big_data @ en.wikipedia.org, n.d.), Big Data pode ser definido como: "Big data é grande Volume, grande Velocidade e/ou grande Variedade de ativos de informação que requerem novas formas de processamento para possibilitar melhorias na tomada de decisão, descoberta de novas perspectivas e otimização de processos.", sendo então popularmente definido por 3V's. Ainda de acordo com o Wikipedia (Big_data @ en.wikipedia.org, n.d.), várias organizações estão adicionando novos Vs como "Veracidade" e "Valor" para complementar o conceito, gerando então a associação de 5V's à definição de Big Data (Ishwarappa & Anuradha, 2015). A figura 1 apresenta esse paradigma.

Figura 1 - Os 5 V's de Big Data

Fonte: elaborado pelo autor

Abaixo, é apresentada a definição dos cinco V's (Cuzzocrea, Saccà, & Ullman, 2013)(Sagiroglu & Sinanc, 2013)(Ishwarappa & Anuradha, 2015) :

Volume: Muitos fatores contribuem para o aumento do volume de dados, como transações armazenadas ao longo dos anos, *streaming* de textos nas mídias sociais, aumento da quantidade de dados coletados por sensores, e outros. No passado o volume excessivo de dados criou um problema de armazenamento, mas com os atuais custos de armazenamento decrescentes, surgem outras questões como por exemplo determinar a relevância entre os grandes volumes de dados e como criar valor a partir destes.

Velocidade: Velocidade significa tanto o quão rápido os dados estão sendo produzidos, quanto o quão rápido os dados devem ser tratados para atender à demanda. Etiquetas RFID, sensores, medidores, uso da internet e mídias sociais estão impulsionando uma necessidade crescente de lidar com torrentes de dados em

tempo quase real. Reagir rápido o suficiente para lidar com a velocidade é um desafio para a maioria das organizações.

Variedade: Os dados de hoje vêm das mais variadas fontes como conversas nas mídias sociais, arquivos de mídia compartilhados nas redes sociais, transações online, utilização de smartphones, dados coletados por sensores e medidores, transações financeiras, e nos mais variados formatos, sejam dados armazenados em bancos de dados tradicionais, como também texto, e-mail, imagem, vídeo, áudio e outros.

Veracidade: Ao longo dos anos, os problemas de qualidade de dados foram identificados como usualmente gerados por erros humanos na entrada de dados e/ou por falta de informações. Na era Big Data, onde a maior parte dos processos de captura de dados é automatizada, problemas similares de qualidade de dados podem ocorrer, devido a mal funcionamento de sistemas ou sensores. Independentemente das razões, deve-se ter cuidado em tratar as inconsistências de dados em Big Data antes de utilizá-los para qualquer tipo de análise.

Valor: Valor é um dos aspectos mais importantes em Big Data, refere-se a utilidade que pode ser extraída dos dados, em outras palavras, a relevância do conhecimento obtido.

2.2 DATA ANALYTICS

O termo Data Analytics pode ser definido como o processo de extração de conhecimento de enormes conjuntos de dados. O processo de Data Analytics requer serviços analíticos sofisticados e escaláveis, bem como ferramentas de programação e aplicações especiais.

Data Analytics aplica-se tanto à área pública, podendo-se citar Cidades Inteligentes como uma importante aplicação, como à área corporativa, aos mais diversos segmentos. Alguns exemplos de aplicação de Data Analytics na área de cidades inteligentes são, em Saúde na análise e tratamento de doença crônicas, em Segurança Pública na gestão de desastres naturais, em Mobilidade Urbana na melhoria de transporte e análise de comportamento, e muitas outras aplicações. Na área corporativa, as aplicações também são inúmeras, podendo-se citar como exemplos, em Varejo para entendimento dos clientes e campanhas de vendas, em Seguros para melhoria do processo de análise de risco no ramo de automóveis, em

Bancos para identificação de Fraudes financeiras, na indústria Automotiva para melhoria de produtos, e muitos outros. (Agrawal, Das, & El Abbadi, 2011).

2.3 DADO ESTRUTURADO E DADO NÃO-ESTRUTURADO

Os dados provenientes das várias fontes acima citadas, nos mais diversos formatos são classificados em dois grandes tipos, dados estruturados e dados semi-estruturados / não-estruturados. Baseando-se em (Bhatia & Vaswani, 2013), abaixo são apresentados os conceitos associados a estes dois grandes tipos.

Dados estruturados tipicamente referem-se a dados tradicionais, que podem ser organizados em tabelas, com linhas e colunas claramente definidas, e armazenados em bancos de dados tradicionais.

Já os dados semi estruturados / não-estruturados são atualmente, em sua maior parte, gerados por “dispositivos” de várias fontes, tais como imagens de satélite, vídeos de sistemas de segurança, sensores, monitoração de clima, como também mídia social, telefonia móvel e outros conteúdos web. Dados vindos dessas fontes muitas vezes não possuem uma organização clara, referem-se a textos, imagens, vídeo, logs web e outros formatos customizados, são um fenômeno recente e os bancos de dados tradicionais não possuem as capacidades requeridas para processar este tipo de informação.

Os dados semi-estruturados / não-estruturados usualmente demandam grande capacidade de armazenamento, são mais difíceis de serem analisados, quando comparados com os dados estruturados, e necessitam ser combinados com dados estruturados para que se tenha uma análise completa.

2.3.1 A Importância e os desafios da análise do dado não-estruturado

Segundo os vários órgãos de pesquisa, em uma organização somente 10% dos dados Big Data são estruturados, e os 90% restantes são não-estruturados (Raza & Xujian, 2020). Apesar dessa grande representatividade, os dados não-estruturados são extremamente subutilizados devido a diversas limitações.

Pesquisa realizada pela Dataversity.net (“DATAVERSITY - Learn Data Management at DATAVERSITY,” n.d.) com 400 empresas de vários setores da economia, indicou alguns fatos importantes referentes ao dado não-estruturado:

- Organizações com capacidade de capturar e gerenciar com sucesso seus dados não-estruturados têm uma vantagem competitiva sobre as organizações que não o são capazes;
- Cerca de 90% das organizações estão planejando ou já estão tratando dados não-estruturados;
- Dificuldades de integração com os sistemas corporativos e imaturidade das ferramentas atualmente disponíveis para tratar dado não-estruturado são as duas principais barreiras para o gerenciamento com sucesso dos dados não-estruturados;
- O volume de dados não-estruturados tende a crescer exponencialmente

A relevância e seu crescimento são os fatores principais que estão motivando os investimentos das organizações em Big Data, em particular no dado não-estruturado.

A análise de dado não-estruturado é normalmente referenciada como o processo de análise do texto não-estruturado, extraíndo informação relevante e transformando-a em informação estruturada, que poderá ser utilizada nas mais diferentes formas (Jagadish et al., 2014) (Bergamaschi, 2014). Estes processos utilizam-se de técnicas como linguística computacional, metadados, taxonomia, data mining, aprendizado de máquina, e outras.

Em um cenário de Big Data, as aplicações de análise de dados não-estruturados são muitas, indo de análise de mídia social, análise de reputação e sentimento, à previsão de churn e fraudes no mercado financeiro, e também aplicações ligadas à área de cidades inteligentes, como Saúde, Gestão de Desastres Naturais, Mobilidade Urbana e outras. É crescente o número de organizações, dos vários setores, que estão buscando formas de combinar dados estruturados e não-estruturados para obter uma visão completa.

2.4 INTERNET DAS COISAS

O número e a diversidade de sensores e dispositivos disponíveis está crescendo fortemente em função da redução do custo da comunicação sem fio, da expansão da Internet, e da redução de tamanho e custos dos microprocessadores (Jara, Bocchi, & Genoud, 2014).

Estes sensores e dispositivos, também chamados de objetos inteligentes, são elementos físicos com a capacidade de serem identificáveis, e opcionalmente, de comunicarem, medirem, detectarem e interagirem com o ambiente e outros objetos inteligentes.

Alguns exemplos de objetos inteligentes são relógios, dispositivos clínicos, sensores de edifícios automatizados, sistemas de controle de acesso, smartphones, TV's inteligentes (as chamadas Smart TV's), e até carros, ônibus e outros meios de transporte.

A evolução do acesso a Internet por meio dos recursos de infraestrutura de comunicação de redes de telefonia móvel e redes de sensores sem fio, está tornando estes objetos inteligentes mais ubíquos nos diferentes ambientes e mais acessíveis pelos sistemas e usuários, formando a chamada Internet das Coisas, ou usando o termo em inglês, Internet of Things (IoT).

Estes dispositivos destinam-se a melhorar qualidade de vida, segurança e desempenho, utilizando-se do potencial da interconexão entre informações, objetos e pessoas.

Traçando uma linha evolutiva, a Internet iniciou com interações do tipo pessoa-pessoa, uma vez que o conteúdo era definido por pessoas, para ser consumido por pessoas. Com a IoT, o conteúdo passou a ser definido por objetos para ser consumido por máquinas, em uma relação máquina-máquina.

Atualmente, o grande desafio é definir e entender a interação entre objetos inteligentes e as pessoas, e como IoT irá representar um papel chave nos ambientes de cidades inteligentes.

2.5 CIDADES INTELIGENTES E MOBILIDADE URBANA

A evolução da humanidade tem caminhado cada vez mais para a vida urbana. Até 2050, 68% da população mundial viverá em áreas urbanas (Nations et al., 2018). Nesse cenário, a demanda por soluções de cidades inteligentes é crescente, em resposta aos desafios atuais da vida urbana, e a mobilidade humana urbana é uma importante área de pesquisa nesse cenário. Um dos principais impulsionadores das soluções de cidades inteligentes, principalmente na mobilidade humana urbana, é a análise de dados de telefones celulares.

O crescimento da população traz aumento de demanda por infraestrutura nas cidades, infraestrutura esta composta por fornecimento de água, energia, transporte, saúde, educação, segurança e outros. Ao mesmo tempo, gera impactos no meio-ambiente como aumento de poluição, enchentes e deslizamentos.

Por outro lado, os avanços e convergência das tecnologias de informação e comunicação têm revolucionado o modo de vida (Naphade et al., 2011) (“International Telecommunication Union (ITU) Statistics,” n.d.):

- Mais de quatro bilhões de pessoas usam a Internet;
- Há mais de oito bilhões de assinantes de telefonia móvel;
- Há trinta bilhões de etiquetas RFID; e
- Os computadores ficam cada vez mais rápidos, menores e mais baratos

Este cenário facilita o controle inteligente de sistemas e torna as pessoas cada vez mais informadas e com maior poder. Neste contexto, uma definição simples de cidade inteligente é: “Integração de tecnologia em uma abordagem estratégica para sustentabilidade, bem-estar do cidadão e desenvolvimento econômico” (Vilajosana et al., 2013).

O número já alto e crescente de assinantes de telefonia móvel faz com que não se tenha dúvida de que a tecnologia de comunicação pessoal móvel é uma das inovações de maior sucesso dos últimos tempos (“International Telecommunication Union (ITU) Statistics,” n.d.) (Naboulsi et al., 2016). Um número crescente de pessoas depende totalmente de seus dispositivos móveis, não só para o trabalho, mas também para a vida pessoal. Uma consequência desse sucesso é que assinantes de telefonia móvel representam hoje uma grande parte da população mundial, onde seus dispositivos móveis estão sempre interagindo com a rede de telecomunicação,

gerando tráfego e eventos georreferenciados. Além disso, o advento dos aplicativos para aparelhos móveis, vem somar uma grande variedade de informações adicionais que podem ser obtidas desses dispositivos.

Com isso, a telefonia móvel é hoje uma fonte rica de geração de dados de grande percentual da população, a um custo mínimo, representando uma importante oportunidade de extração de conhecimento na área de mobilidade urbana, tanto em transporte inteligente com em dinâmica humana.

Dessa forma, a evolução de IoT somada à capacidade de análise de Big Data são importantes componentes para a viabilização da transformação das cidades em cidades mais inteligentes. Pode-se dizer que o triângulo formado por IoT, Big Data e mobilidade urbana definem um ecossistema que possibilita maior conhecimento sobre a dinâmica e comportamento da sociedade.

2.6 ANÁLISE DE DADOS E MOBILIDADE URBANA: AS OPORTUNIDADES

Uma cidade inteligente integra e otimiza um conjunto de sistemas públicos e privados, buscando atingir níveis melhores de eficiência e eficácia na gestão dos recursos, trazendo benefícios ao meio ambiente e à sociedade (Naphade et al., 2011).

Estes sistemas correspondem a Saúde, Transportes, Gestão de Águas e Energia, Segurança, Educação, Gestão de Desastres e outros, são tanto consumidores de informação externas, como também geradores de informações consumidas entre eles e externamente, formando um grande sistema de sistemas interligados.

A transformação requerida por mobilidade urbana em cidades inteligentes depende do uso de técnicas de análise avançadas que possibilitem extrair conhecimento dos dados gerados por eventos do mundo real, buscando melhorar os processos urbanos. A Figura 2 apresenta uma ilustração deste sistema de sistemas que suportam as iniciativas de cidades inteligentes:

Figura 2 - Cidade inteligentes: Um sistema de sistemas

Fonte: Elaborado pelo autor

O conhecimento gerado pela análise de dados proveniente dos sensores e medidores instalados nas cidades, dos smartphones utilizados pelos indivíduos, dos dados provenientes dos sistemas e outros, gera oportunidades de aplicações nas diversas áreas acima mencionadas, proporcionando um ciclo contínuo de melhorias.

No contexto de mobilidade urbana em cidades inteligentes, a análise de dados pode ser utilizada nas três grandes etapas que compõem o macroprocesso de gestão de uma cidade (Naphade et al., 2011), a seguir apresentadas:

Planejamento: Nesta etapa é realizado o planejamento dos recursos. Para tal, a análise de dados contribui com fontes de informações históricas, dados que permitem entender o comportamento dos cidadãos, reconhecimento de padrões, recursos para comparação com outras cidades potencializando o compartilhamento de lições aprendidas e melhores práticas, análise de tendências para previsão de demandas futuras, e outros. Citando alguns exemplos: o planejamento aplica-se à Saúde, no melhor entendimento das doenças crônicas e planejamento das medidas preventivas e de acompanhamento; em Transportes na melhor compreensão dos fluxos de mobilidade urbana possibilitando o planejamento dos recursos de transporte adequados; na Educação com o entendimento da demanda, oferta e resultados obtidos, escalonados por região, renda, idade, e outros possibilitando melhor planejamento dos recursos a serem oferecidos e medidas de gestão.

Gerenciamento: Esta etapa corresponde à gestão dos recursos, onde a análise de dados propicia uma visão integrada dos diferentes recursos e seu uso, bem como uma melhor gestão da manutenção e evolução destes. Por exemplo, uma ação de manutenção na rede elétrica pode ser combinada com uma manutenção preventiva da rede de telefonia, em data/horário que gere o menor impactos aos usuários, baseado no padrão de uso e fluxo na área.

Operação: Nesta etapa é realizada a operação propriamente dita dos recursos. O tratamento dos dados recebidos em tempo real provenientes dos diversos sensores, medidores, smartphones geram conhecimento essencial para a visão de uso dos recursos, previsão de eventos e de demanda do futuro próximo. Por exemplo, a concessionária de distribuição de energia pode, através de modelos sofisticados, combinar informações sobre o clima em tempo real com dados históricos para prever problemas em sua rede, tomando então ações proativas.

Estas três etapas se retroalimentam, gerando um ciclo de melhorias contínuas.

A Figura 3 sumariza o macroprocesso e suas três etapas:

Figura 3 - Etapas do macro-processo de cidades inteligentes

Fonte: Elaborado pelo autor

Traçando um paralelo do macroprocesso acima para uma visão de tratamento e análise das informações, pode-se representar o ciclo contínuo de melhorias por meio da Figura 4 (Naphade et al., 2011) (Vilajosana et al., 2013):

Figura 4 - Ciclo contínuo de melhorias

Fonte: Elaborado pelo autor baseado nas referências citadas

As possibilidades de aplicação de análise de dados em cidades inteligentes na área de mobilidade urbana são inúmeras.

2.7 ANÁLISE DE DADOS E MOBILIDADE URBANA: OS DESAFIOS

Neste item são apresentados os principais desafios pelos quais passam as iniciativas de aplicações de análise de dados na área de mobilidade urbana. Para tal, é tomado como base o processo típico, do ponto de vista de tratamento das informações, associado à implementação de uma iniciativa de análise de dados em mobilidade urbana, apresentando os principais desafios associados a cada uma das etapas que compõem o processo.

2.7.1 O processo de análise de dados

O processo de análise de dados ou Data Analytics, mencionando o termo em inglês, é complexo e composto de diversas etapas, requer métodos e técnicas específicas, que são tratadas por perfis de usuário distintos.

Enfatizando, o foco desta pesquisa será restrito aos dados provenientes de operadoras de telefonia móvel, referentes ao uso da telefonia pelos assinantes.

As principais etapas que compõem um processo típico de análise de dados no contexto de mobilidade urbana, quer seja em tempo real ou não, são representadas na Figura 5, baseada em (Talia, 2013) e (Khan, Anjum, & Kiani, 2013):

Figura 5 - Ciclo completo de análise de dados

Fonte: Figura gerada pelo autor a partir das referências acima mencionadas

Onde:

Coleta: Diferentemente da coleta tradicional de dados, na análise de dados em mobilidade urbana os dados são provenientes das mais diversas fontes, são de variados tipos, podendo ser estruturados, semi-estruturados ou não estruturados. Esta etapa de aquisição tem os “4V’s” de big data como importante desafio: grande **V**olume, alta **V**elocidade, grande **V**ariiedade e **V**eracidade. No contexto de mobilidade urbana, muitos dos dados provêm de sensores, medidores, smartphones e outros dispositivos associados a IoT, gerando grande complexidade a esta etapa.

Armazenamento: Com o avanço das tecnologias, o armazenamento dos diversos tipos e grandes volumes de dados em bases de dados adequadas a cada tipo, acompanhados de seus metadados não é mais uma dificuldade atualmente. Por outro lado, garantir a segurança e privacidade dos dados, que muitas vezes estão distribuídos, frequentemente utilizam ambientes de computação em nuvem, ou são provenientes de operadoras de telefonia móvel e contêm identificação pessoal, representa um grande desafio.

Tratamento: Corresponde à etapa de tratamento de qualidade nos dados, contemplando limpeza, enriquecimento, aplicação de filtros para seleção dos dados que atendam aos critérios de veracidade definidos no processo, interligação e agregação dos dados, gerando uma massa de dados consistente para a realização das análises. Atenção especial deve ser dada aos critérios de seleção, qualidade,

enriquecimento e veracidade para garantir que estes sejam adequados aos objetivos de análise, bem como a anonimização de dados pessoais sempre que necessário para preservação de privacidade.

Processamento: Esta etapa consiste dos tratamentos preparatórios dos dados necessários para as análises desejadas, podendo corresponder tanto a simples consolidações como geração de séries temporais, e outros. Novamente, devem ser tomados os cuidados necessários para proteção das informações sensíveis.

Análise: Corresponde à análise dos dados obtidos, efetuando-se refinamentos, agregando-se novos dados, executando-se modelos estatísticos e de aprendizado de máquina, até chegar-se aos resultados desejados, gerando-se conhecimento a partir de dados coletados.

Visualização: Nesta etapa ocorre a visualização dos processamentos e análises efetuadas, sendo frequentemente utilizada de modo interativo com a etapa de Análise, até que se chegue ao resultado final.

Ações: A partir dos resultados obtidos nas análises definem-se ações a serem implementadas, podendo estas serem tanto de execução imediata, como de médio / longo prazo, envolvendo um ou mais sistemas, gerando interação direta com os indivíduos.

O processo como um todo é altamente iterativo e recursivo, algumas vezes retornando à etapa de coleta para buscar dados que sejam necessários para complementar a análise, até alcançar a etapa final, onde é obtido o conhecimento necessário para a execução de ações.

2.7.2 Os desafios para obtenção dos resultados desejados

Cada uma das etapas do processo de análise de dados em mobilidade urbana apresenta diferentes desafios para a obtenção dos resultados desejados. O quadro a seguir apresenta um sumário preliminar destes desafios, que foram aprofundados durante a realização da pesquisa.

Quadro 1- Desafios de análise de dados em mobilidade urbana

Etapa do Processo	Principais desafios em Análise de Dados & Mobilidade Urbana
Coleta	<ul style="list-style-type: none"> – Acordos para obtenção dos dados – Custo x Benefício da coleta de dados – Segurança e Privacidade – Variedade e variabilidade das fontes de informação – Velocidade e Periodicidade da coleta de dados – Sincronismo de tempo nas aplicações em tempo real
Armazenamento	<ul style="list-style-type: none"> – Garantia da segurança física e lógica dos dados armazenados – Proteção dos dados no armazenamento
Tratamento e Processamento	<ul style="list-style-type: none"> – Seleção dos dados pertinentes às análises desejadas – Qualidade dos dados – Modelos para integração das informações – Segurança e Privacidade – Controle de acesso – Padrões e Interoperabilidade – Tempo de processamento
Análise e Visualização	<ul style="list-style-type: none"> – Modelos comprovados para a análise dos dados – Dados para calibrar e treinar os modelos com acuracidade – Técnicas avançadas de <i>analytics</i> – Tempo de atualização da análise e visualização – Validação das descobertas extraídas da análise dos dados
Ações	<ul style="list-style-type: none"> – Controle otimizado dos vários sistemas relacionados às ações – Prazo para acionamento – Interação com os cidadãos

Fonte: Elaborado pelo autor

2.8 TELEFONIA MÓVEL COMO FONTE DE DADOS

O uso de telefones celulares pela população mundial tem aumentado a cada ano, tanto nos países desenvolvidos como nos em desenvolvimento ("International Telecommunication Union (ITU) Statistics," n.d.). Com isso, tem-se a geração de uma grande variedade de dados provenientes dos usuários de telefones celulares que trazem um rico conhecimento sobre localização e mobilidade dos indivíduos, auxiliando no entendimento da dinâmica humana nos centros urbanos.

Os principais conjuntos de dados produzidos pelos usuários de telefones celulares são os registros de uso das redes de telefonia móvel, os gerados pelas mídias sociais como Twitter por exemplo, os dados originados pelos aplicativos instalados nos aparelhos, e os registros de posicionamento gerados pelos próprios telefones móveis (Celes et al., 2019), cada um deles com vantagens e limitações.

Nesse contexto, os registros de uso da rede de telefonia móvel emergem como uma fonte de dados valiosa, amplamente utilizada em obras de mobilidade urbana, apresentando diversas vantagens sobre as demais, com cobertura de grandes áreas urbanas, grande quantidade de usuários com alta taxa de penetração na população dos mais diferentes países, e sem necessidade de infraestrutura ou custos adicionais para a coleta de dados (Celes et al., 2019) ("International Telecommunication Union (ITU) Statistics," n.d.).

Os registros de uso da rede de telefonia móvel contêm informações espaço temporal de cada chamada de voz, mensagens de texto, uso de dados e atividade na Internet de cada um dos assinantes atendidos.

Esses registros podem ser basicamente de dois tipos (Calabrese et al., 2014) (Zhang, Fu, Li, & Chen, 2019):

- Registros detalhados da chamada (CDRs): Um CDR, do inglês Call Detail Record, contém os detalhes de uma chamada telefônica de voz ou troca de mensagem de dados. Em geral, um CDR consiste do número de telefone de origem e de destino, data e hora de início da chamada, duração, o tipo de comunicação (chamada de voz ou dados), identificação e localização da antena ou torre de celular envolvida na chamada.
- Registros de detalhes do protocolo da Internet (IPDRs): um IPDR, do inglês Internet Protocol Data Record, contém detalhes do uso da Internet. Normalmente, contém a

identificação do telefone celular, data e hora, número de bytes transferidos, site visitado e identificação da antena à qual o telefone se conecta.

3 ANÁLISE DE DADOS DE MOBILIDADE URBANA – MÉTODO DE TRABALHO

O desenvolvimento da pesquisa está baseado no método de Estudo de Caso(Yin, 2003).

Para a realização da pesquisa foram selecionadas diversas referências de aplicação de análise de dados em mobilidade urbana, com foco no uso de fontes de dados proveniente de dispositivos pessoais móveis. A partir das observações dessas referências em relação à forma de uso dos dados de telefonia móvel, métodos e técnicas adotados, resultados obtidos e dificuldades encontradas, foi compilado um conjunto de recomendações a ser aplicado na análise de dados em mobilidade urbana referente às áreas selecionadas.

A seguir é apresentado o método utilizado para o desenvolvimento da pesquisa.

3.1 VISÃO GERAL

O estudo de caso considera referências na área de análise de dados provenientes da telefonia móvel, em diferentes aplicações, todas em mobilidade urbana. Objetiva ter-se um estudo que contemple referências com características variadas dando maior embasamento aos resultados obtidos.

Por referências entendem-se artigos de periódicos e conferências, dissertações e teses, consultas a livros e informações disponíveis no meio acadêmico ou públicas.

As áreas de aplicação consideradas têm como foco o tratamento dos dados provenientes do uso da rede de telefonia móvel, no formato de dados agregados, isto é, aplicações que não necessitem de informações referentes a cada indivíduo para realização das análises.

Nas referências estudadas são observados os modelos, métodos e técnicas usados para o tratamento dos dados provenientes do uso da rede de telefonia móvel, observando-se o tipo de aplicação, dificuldades encontradas e resultados obtidos.

Para apoiar o estudo de caso, foi definido um modelo para classificação e comparação dos elementos observados, considerando as etapas do ciclo de análise de dados.

3.2 ETAPAS DO DESENVOLVIMENTO DA PESQUISA

Para o desenvolvimento da pesquisa foram percorridas cinco etapas, iniciando-se com a Etapa 1, onde é efetuada a definição dos critérios para seleção das referências e a seleção propriamente dita, na sequência passa-se para a Etapa 2, na qual é definido o modelo de observação, continuando com a Etapa 3, de observação das referências selecionadas segundo o modelo definido, seguida pela Etapa 4, quando é realizada a compilação dos resultados obtidos com a observação das referências, finalizando-se com a Etapa 5, que consiste na elaboração das recomendações positivas e negativas a partir da avaliação efetuada nas etapas anteriores. A cinco etapas mencionadas estão apresentadas de forma sumarizada na Figura 6 e são descritas em maiores detalhes nas subseções seguintes.

Figura 6 - Etapas do desenvolvimento da pesquisa

Etapa	1	2	3	4	5
	Seleção das referências	Modelo de observação	Observação das referências	Avaliação dos resultados	Recomendações
Sumário	Definição de critérios, e seleção das referências para o estudo	Definição do modelo de observação das referências	Observação das referências de acordo com o modelo definido	Compilação dos resultados obtidos com a observação das referências	Preparação das recomendações positivas e negativas

Fonte: Elaborado pelo autor

3.2.1 Etapa 1 – Seleção das referências

A primeira etapa do desenvolvimento da pesquisa consistiu em selecionar as referências que fazem parte do estudo. Esta etapa é de extrema importância para os resultados do trabalho, onde as referências selecionadas devem ser variadas o suficiente para que os resultados obtidos sejam significativos.

Essa etapa iniciou-se com a definição dos critérios para a busca e seleção das referências que fazem parte do estudo, incluindo por exemplo itens como: tipo principal de fonte de dados utilizada, ano de publicação, temporalidade da análise, abrangência e outros.

Uma vez definidos os critérios, foi efetuada pesquisa bibliográfica para busca e separação inicial das referências de análise de dados em mobilidade urbana que atendam aos critérios definidos.

Após a separação inicial, as referências foram verificadas e então selecionadas as que possuíam conteúdo que possibilitava a observação no nível de detalhe necessário para esta pesquisa. O conjunto de referências selecionadas foi o objeto do estudo de caso.

3.2.2 Etapa 2 – Modelo de observação

Para observação das referências, técnicas e métodos empregados, e resultados obtidos, foi primeiramente proposto um modelo de avaliação a ser aplicado ao universo de referências selecionadas. Este modelo de avaliação compreende uma forma de classificação dos elementos que compõem as referências e a compilação das observações, com objetivo de explorar como os dados de uso da telefonia móvel utilizados nas aplicações de mobilidade urbana.

3.2.3 Etapa 3 – Observação das referências

Tendo como base o modelo definido na Etapa 2, foi realizada a observação das referências selecionadas, avaliando-se o tipo de aplicação, fontes de dados utilizadas, métodos e técnicas empregados.

Durante essa observação, foram também identificados os possíveis agrupamentos, as características de cada um deles, e efetuada a sumarização dos métodos e técnicas usados.

3.2.4 Etapa 4 – Avaliação dos resultados

Nesta etapa foram avaliados os resultados obtidos na observação das referências, visando a preparação das recomendações.

A avaliação dos resultados obtidos é composta da compilação e consolidação da observação efetuada na etapa 3, contemplando diferenças das características dos agrupamentos identificados, se aplicável, pontos relevantes sobre os diferentes métodos usados, desafios percebidos e resultados obtidos.

3.2.5 Etapa 5 – Recomendações

A partir da observação das referências e da avaliação dos resultados, realizados nas etapas 3 e 4 respectivamente, foram geradas as recomendações positivas e negativas, que têm por objetivo traçar um panorama dos métodos que trouxeram os resultados esperados na análise de dados em mobilidade urbana, e dos pontos que trouxeram dificuldades.

Estas recomendações poderão servir de subsídios para casos futuros, acelerando o processo de obtenção de resultados.

4 ANÁLISE DE DADOS DE MOBILIDADE URBANA – DEFINIÇÃO DE UM FRAMEWORK ANALÍTICO

Este capítulo descreve os resultados obtidos pela pesquisa realizada, bem como apresenta a definição de um framework analítico a ser utilizado para análise de dados de mobilidade urbana.

São apresentadas as referências selecionadas para o estudo de caso objeto deste trabalho, relatando como os dados de telefonia móvel são aplicados à observação espaço temporal da mobilidade humana urbana, explorando as fontes de dados utilizadas, técnicas e métodos aplicados, e resultados obtidos.

A exploração das referências se concentra nas fontes de dados, áreas de aplicação e resultados alcançados, uma vez que o foco desta pesquisa é discutir o potencial do uso de dados agregados de telefonia móvel em aplicações de mobilidade humana urbana. São também observados os desafios relatados.

A organização deste capítulo obedece a sequência de cinco etapas descritas no capítulo 3.

4.1 SELEÇÃO DAS REFERÊNCIAS

A seleção das referências para o estudo de caso foi realizada em dois passos, onde no primeiro passo foi feita a busca nas bibliotecas acadêmicas de publicações segundo os critérios definidos, e no segundo passo foi efetuada a seleção propriamente dita das referências que foram objeto do estudo de caso.

Esses dois passos estão descritos a seguir.

Passo 1: Definição dos critérios de busca e seleção inicial

Conforme descrito na metodologia, o primeiro passo para seleção das referências é a definição dos critérios de pesquisa.

Os critérios adotados para o primeiro filtro foram os seguintes: aplicação em mobilidade urbana, dados de telefonia móvel, ano de publicação a partir de 2015, publicações nas áreas de tecnologia, como engenharia das diversas modalidades, ciência da computação, ciência da informação e estudos urbanos, e filtros adicionais

de indexação selecionados caso a caso, tais como mineração de dados, análise de dados, smartphones, e outros.

Foram acessadas as principais bibliotecas de publicações acadêmicas, listadas no Quadro 2, e efetuada a busca segundo os critérios definidos. Para cada lista de referências obtidas como resultado da pesquisa de cada uma das bibliotecas, foram avaliadas as 50 primeiras por ordem de relevância, verificando-se primeiramente os títulos, seguidos pela leitura breve dos sumários, excluindo-se as referências não pertencentes ao escopo da pesquisa. Esse novo filtro resultou em um subconjunto de 149 referências.

Quadro 2 - Bibliotecas digitais

Biblioteca	Link
ACM Digital Library	https://dl.acm.org/
Scopus	https://www.scopus.com
IEEE Xplore	https://ieeexplore.ieee.org/Xplore/home.jsp
Springer	https://www.springeropen.com/
Elsevier	https://www.sciencedirect.com/
Web of Science	http://apps.webofknowledge.com
Portal de Periódicos CAPES/MEC	http://www.periodicos.capes.gov.br/

Fonte: Elaborado pelo autor

Passo 2: Seleção das referências para o estudo de caso

Para o subconjunto de 149 referências obtido no passo 1, foi efetuada a verificação do conteúdo de cada documento observando-se os seguintes critérios: trabalho referente à aplicação de análise de dados em mobilidade urbana, fonte principal de dados proveniente do uso da rede de telefonia móvel, tratamento dos dados de telefonia móvel de forma agregada, conteúdo do documento incluindo descrição dos conjuntos de dados utilizados, métodos aplicados e resultados alcançados.

Neste passo procurou-se obter um subconjunto de 15 a 20 referências em temas variados, limitando a fonte principal de dados, conforme objetivo desta pesquisa, de forma a ter-se uma amostra com boa representatividade para a observação do potencial da análise dos dados agregados do uso de telefonia móvel na área de mobilidade urbana.

Os dados provenientes da telefonia móvel proporcionam diversas aplicações na área de mobilidade urbana, onde as fontes geradoras podem ser tanto os registros

referentes ao uso da rede de telefonia móvel, como os produzidos pelo uso de redes sociais, como também os relacionados aos aplicativos de serviços de mobilidade. Assim, durante o processo de verificação dos documentos selecionados no primeiro passo foram encontrados 61 documentos que utilizam conjuntos de dados referentes ao uso da rede de telefonia móvel e 88 documentos que utilizam outros tipos de dados como fonte principal. Como exemplos desses que utilizam outros tipos de dados como fonte principal podem ser citados o uso de dados provenientes da rede social Twitter (Fiorio et al., 2017), da rede social Foursquare (Kadar & Pletikosa, 2018), e de GPS de taxis (Yuan et al., 2015).

Voltando aos 61 documentos encontrados que utilizam conjuntos de dados referentes ao uso de telefonia móvel, verificou-se que 33 deles fazem uso de registros individuais, em geral relacionados à análise de origem e destino das trajetórias dos indivíduos para planejamento de transporte (Thuillier, Moalic, Lamrous, & Caminada, 2018) (Graells-Garrido, Caro, & Parra, 2018), análise de comportamentos individuais para fins de estudos demográficos (Jahani et al., 2017), e outros.

Dos 28 documentos restantes, foram então filtrados os que apresentam experimentos que fazem uso dos dados de uso da rede de telefonia móvel de forma agregada e possuem informações suficientes para a observação.

Este segundo passo resultou então na seleção de 17 referências para o estudo de caso desta pesquisa. Os trabalhos selecionados encontram-se relacionados no Quadro 3 e resumidamente apresentados a seguir.

Exploração dos dados de telefonia móvel para extrair informações sobre a dinâmica humana, combinando-as com dados de uso das áreas urbanas e informações sociodemográficas, com o objetivo de testar as quatro condições que promovem “vida” em uma cidade segundo a teoria de Jacobs (De Nadai et al., 2016). O estudo é aplicado em seis grandes cidades italianas, Bologna, Florença, Milão, Palermo, Roma e Turin.

Avaliação do potencial da análise de registros de chamadas de telefonia móvel para detectar padrões de uso das áreas urbanas (Ríos & Muñoz, 2017). Com o método proposto detectam-se quatro perfis de uso das áreas urbanas: residencial, trabalho, lazer & comércio, e horas de pico. Para teste de seu modelo utiliza dados de telefonia móvel da cidade de Santiago no Chile.

Quadro 3 - Lista de referências selecionadas

Título	Referência
Death and Life of Great Italian Cities	(De Nadai et al., 2016)
Land Use detection with cell phone data using topic models: Case Santiago, Chile	(Ríos & Muñoz, 2017)
Exploiting mobile phone data for multi-category land use classification in Africa	(Mao, Thakur, & Bhaduri, 2016)
Urban Electric Load Forecasting with Mobile Phone Location Data	(Selvarajoo et al., 2018)
Energy consumption prediction using people dynamics derived from cellular network data	(Bogomolov et al., 2016)
Using Mobile Phone Data for Electricity Infrastructure Planning	(Martinez-Cesena, Mancarella, Ndiaye, & Schlöpfer, 2015)
Crowd Management: A New Challenge for Urban Big Data Analytics	(Celes et al., 2019)
Moves on the Street: Classifying Crime Hotspots Using Aggregated Anonymized Data on People Dynamics	(Bogomolov et al., 2015)
Countrywide arrhythmia: emergency event detection using mobile phone data	(Gundogdu, Incel, Salah, & Lepri, 2016)
Research of urban land use and regional functions based on mobile data traffic	(Miao, Qiao, & Yang, 2018)
Inferring Land Use Type in Urban Area with Mobile Big Data	(Xing et al., 2018)
Fusing GPS Probe and Mobile Phone Data for Enhanced Land-use Detection	(Furno, Faouzi, Fiore, & Stanica, 2017)
Improving land use inference by factorizing mobile phone call activity matrix	(Mao, Ahn, Bhaduri, & Thakur, 2017)
Understanding Urban Dynamics From Massive Mobile Traffic Data	(Zhang et al., 2019)
A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas	(Furno, Fiore, Stanica, Ziemlicki, & Smoreda, 2017)
Comparing and modelling land use organization in cities	(Lenormand et al., 2015)
Electricity Demand and Population Dynamics Prediction from Mobile Phone Metadata	(Wheatman, Noriega, & Pentland, 2016)

Fonte: Elaborado pelo autor

Apresentação de um método para inferir o tipo de uso das áreas urbanas na cidade de Dakar no Senegal, África (Mao et al., 2016). Em países de baixa renda, as pesquisas sobre o uso das áreas urbana são raras, assim, a abordagem apresentada cria uma alternativa para obtenção desse conhecimento. Classificam as áreas urbanas em dez categorias, sendo elas residencial, comercial-industrial /

escritório, comercial- negócios / varejo / lazer, comercial de alta e baixa densidade, residencial de alta e baixa densidade, áreas de uso misto como centros comerciais e residenciais da cidade.

Proposição de um novo método de previsão de demanda de eletricidade usando dados de telefonia móvel (Selvarajoo et al., 2018). As características extraídas dos registros provenientes do uso de telefonia móvel descrevem a diversidade e regularidade da dinâmica humana em cada localidade, possibilitando mapear a variação de distribuição da população ao longo do tempo. Para demonstração do método proposto utiliza dados de telefonia móvel da região de Trentino, Itália.

Introdução de uma nova abordagem para prever o consumo de energia elétrica da próxima semana baseado na análise da dinâmica humana derivada dos registros de chamadas de telefonia móvel (Bogomolov et al., 2016). Essa abordagem tem dois objetivos principais, o primeiro é a previsão do consumo médio diário de energia nos próximos sete dias, e o segundo é prever o pico diário de consumo de energia, com isso é possível obter-se subsídios para aumentar a eficiência tanto dos geradores como dos distribuidores de energia, e também atender aos picos de demanda dos consumidores. Para validação da abordagem, são usados dados de telefonia móvel da região de Trentino, Itália.

Avaliação da contribuição dos dados de telefonia móvel para o desenvolvimento de modelos de demanda de energia, a fim de aprimorar os estudos de planejamento de energia e as práticas de eletrificação já existentes (Martinez-Cesena et al., 2015). Verificam a adequação dos dados de telefonia móvel como indicador para as necessidades atuais e futuras de consumo de eletricidade. Avalia também o uso de dados de telefonia móvel como indicador de aumento de população em certa região em função de movimentos migratórios. Para demonstração do modelo utiliza dados de telefonia móvel do Senegal como representante de um país em desenvolvimento.

Exposição dos benefícios de utilizar registros de chamadas de telefones móveis combinados com dados provenientes de redes sociais e de Pontos de Interesse (POI) para detectar a ocorrência de situações de aglomerações de pessoas (Celes et al., 2019). Para demonstração do modelo proposto utilizam dados de telefonia móvel da cidade de Milão, Itália.

Apresentação de uma nova abordagem para a previsão de áreas de alto risco de criminalidade a partir de dados do comportamento humano derivados da atividade

de uso da rede de telefonia móvel em combinação com informações demográficas (Bogomolov et al., 2015). Descrevem uma metodologia para prever quando uma área geográfica de uma grande metrópole terá alto ou baixo nível de criminalidade no mês seguinte. No estudo são utilizados dados da cidade de Londres.

Proposição de uma metodologia para, a partir dos dados de registro de uso de telefonia móvel, identificar e caracterizar padrões de chamadas atípicas associados tanto a situações de emergência, como eventos violentos ou tumultos, como também a eventos sociais como jogos ou comemorações de feriados (Gundogdu et al., 2016). Partem da premissa de que as pessoas tendem a fazer ligações quando ocorrem eventos não esperados. Para validação da metodologia proposta utilizam dados de telefonia móvel da República da Costa do Marfim.

Investigação dos principais padrões de comportamento humano online para descobrir as diferentes áreas funcionais nas cidades (Miao et al., 2018). Definem um método para identificação do uso das áreas urbanas a partir de informações obtidas do tráfego de dados móveis e relacionamento deste com as atividades humanas. A técnica é aplicada sobre dados coletados pelas principais operadoras de telefonia móvel da China.

Proposição de uma metodologia para detecção de uso das áreas urbanas com uso registros de chamadas, de texto e de tráfego de móvel, baseada em padrões de assinaturas (Xing et al., 2018). O método é aplicado a um conjunto de dados coletado no nordeste da China.

Investigação dos padrões de uso das áreas a partir do estudo simultâneo de rastros GPS de veículos exploradores com os dados de uso de telefonia móvel (Furno, Faouzi, et al., 2017). A técnica é aplicada em cidades da França e da Itália, e conclui que o uso combinado dos dados de uso da rede de telefonia móvel combinados com rastreamento GPS possibilitam caracterizar o tipo de uso das áreas urbanas com maiores detalhes.

Proposição de analisar o uso das áreas por meio da caracterização dos padrões de atividade humana com base no volume agregado de chamadas de telefonia móvel, e com a aplicação de técnicas específicas identificar classes comportamentais fundamentais (Mao et al., 2017). Para validação do modelo proposto utilizam dados de telefonia móvel de Dakar, Senegal, descobrindo dois padrões fundamentais: comercial / escritórios / industrial e residencial.

Apresentação de uma nova abordagem para compreender os padrões de consumo de dados móveis e, através dessa compreensão, revelar a atividade e ecologia humana nas áreas urbanas (Zhang et al., 2019). O modelo proposto possibilita tanto a detecção de padrões de atividade diária humana como também a detecção de eventos incomuns, como concertos e partidas de futebol. O modelo é aplicado em Xangai, utilizando registros de tráfego móvel da operadora de telefonia.

Investigação dos vários padrões que surgem na atividade humana de comunicação móvel registrados nas regiões metropolitanas (Furno, Fiore, et al., 2017). Para tal, introduzem uma técnica para identificar classes de assinaturas de tráfego móvel, que são distintas nas diferentes áreas urbanas, possibilitando assim a identificação do uso funcional das áreas com maior nível de detalhe. Aplicam a técnica a extensos volumes de dados coletados pelas principais operadoras de telefonia móvel em 4 cidades da Itália e 6 cidades da França, analisando as similaridades e diferenças entre os padrões encontrados para os dois países.

Proposição de uma abordagem de rede funcional para determinar padrões de uso das áreas a partir dos registros de uso de telefonia móvel (Lenormand et al., 2015). O método proposto é aplicado sobre os dados das cinco mais populosas cidades da Espanha, detectando quatro tipos principais de uso das áreas que correspondem a diferentes padrões temporais. O método permite realizar a comparação entre as cidades espanholas de vários tamanhos, verificando forte semelhança entre as cidades para os tipos principais de uso das áreas, e que os usos mistos são específicos para cada cidade.

Introdução de um método de aprendizado de máquina para previsões diárias do consumo de energia por hora, com base na dinâmica da população e demanda de eletricidade (Wheatman et al., 2016). O método proposto é aplicado sobre os dados de um pequeno país europeu, sendo capaz de prever a demanda de energia com precisão abaixo de 3% de erro.

4.2 MODELO DE OBSERVAÇÃO

Para observação das referências, métodos empregados e resultados obtidos foi proposto um modelo de avaliação aplicado ao universo de referências

selecionadas. O modelo de observação proposto para o estudo de caso compreende uma forma de classificação dos elementos observados.

Este modelo inclui três componentes, sendo que o primeiro corresponde à classificação das áreas de aplicação em mobilidade urbana encontradas no estudo de caso, o segundo à tipificação das formas de contribuição das fontes de dados de telefonia móvel, o terceiro refere-se às técnicas de pré-processamento e análise observadas no estudo de caso. A Figura 7 ilustra o modelo proposto.

Figura 7 - Modelo de Observação

Fonte: Elaborado pelo autor

A observação das referências foi realizada segundo o modelo proposto na Figura 7. São apresentados a seguir os resultados para cada um dos componentes do modelo.

4.2.1 Áreas de aplicação em mobilidade urbana

Na observação das referências foram identificadas três grandes áreas de aplicação da análise de dados agregados de telefonia móvel em mobilidade urbana, sendo a primeira delas o uso funcional das áreas urbanas na área de Planejamento urbano, a segunda corresponde à demanda de energia elétrica na área de Energia, e a terceira Segurança, sendo esta composta por duas subáreas: gestão de multidões e previsão de criminalidade.

A seguir é apresentada a descrição de cada uma dessas áreas, e o Quadro 4 exibe um sumário.

Quadro 4 – Sumário das áreas de aplicação em mobilidade urbana

Classificação da área de aplicação		Descrição sumarizada da contribuição dos dados de telefonia móvel na área de aplicação	Sub-área x Contribuição dos dados de telefonia móvel	
Área	Sub-área		Informação extraída	Uso da informação
Planejamento Urbano	Uso funcional das áreas urbanas	A visão espaço temporal de distribuição da população nas áreas urbanas é utilizada para detectar a dinâmica humana e, através desta, descobrir padrões que são associados ao uso funcional das áreas urbanas	Estimativa de distribuição espaço-temporal da população	Deteção de padrões de ocupação das áreas urbanas
Energia	Consumo de energia	O mapeamento da distribuição da população nos locais ao longo do tempo traz subsídios tanto para previsão de demanda de consumo de energia elétrica como também planejamento de infraestrutura para fornecimento de eletricidade	Estimativa de distribuição espaço-temporal da população	Modelo preditivo de consumo médio de energia elétrica e de pico para as localidades
Segurança	Gestão de multidões	Os padrões de uso da rede de telefonia móvel no espaço e no tempo possibilitam detectar anomalias, associando-as a ocorrência de eventos de aglomerações de pessoas	Estimativa de atividade da população, na visão espaço-temporal	Identificação de anomalias no padrão de atividade nas localidades
	Previsão de criminalidade	A visão da dinâmica espaço temporal da população é correlacionada a ocorrências de crimes, contribuindo para a previsão de nível de criminalidade nas áreas urbanas	Estimativa de distribuição espaço-temporal da população	Modelo preditivo de nível de criminalidade

Fonte: Elaborado pelo autor

Planejamento urbano – Uso funcional das áreas urbanas

Na observação das referências do estudo de caso, o uso da análise de dados agregados de telefonia móvel para a detecção do uso da terra surge como uma das principais áreas de aplicação em mobilidade urbana, como por exemplo em Ríos e Muñoz (2017), Furno et al. (2017) e Xing et al. (2018).

Neste tipo de aplicação, a visão espaço temporal de distribuição da população é utilizada para detectar a dinâmica humana e, através desta, descobrir padrões que são associados ao uso funcional das áreas urbanas.

O uso funcional das áreas urbanas, também denominado de uso da terra, pode ser definido como “a forma de utilização humana da terra” caracterizada pelos agrupamentos, atividades e serviços que as pessoas realizam (Mao et al., 2017). É

importante diferenciar os conceitos de 'cobertura da terra' e 'uso da terra': o primeiro refere-se ao estado biofísico da terra, enquanto o segundo corresponde ao seu uso socioeconômico.

Cada cidade pode ser dividida em diferentes áreas funcionais para diferentes propósitos, conforme as finalidades socioeconômicas, assim, o uso da terra pode ser classificado em categorias, como residencial, escritórios, lazer, comercial ou industrial. Alguns exemplos encontram-se em De Nadai et al. (2016), Lenormand et al. (2015) e Miao et al. (2018).

A classificação correta do tipo de uso funcional das áreas é crucial para entender a estrutura da cidade, podendo esta impulsionar os investimentos, o planejamento sustentável e os processos de tomada de decisão em domínios como urbanismo, transporte e infraestrutura, criando um ambiente eficiente e sustentável para os cidadãos. Ainda, a compreensão do uso da terra é também de grande valia para subsidiar estimativa de população, zoneamento de bairros, além de análises de riscos e poluição (Mao et al., 2017).

As abordagens tradicionais para classificação do uso da terra urbana normalmente utilizam pesquisas de campo baseadas em questionário de amostras populacionais ou dados de imagens de satélites. As realizações dessas pesquisas são caras, demoradas e sujeitas a rápida obsolescência. Já a análise dos dados de imagens de satélites, possibilita a detecção de áreas como urbanas, agrícolas, água, florestas e outras, tendo melhor uso para classificação da cobertura da terra e não do uso funcional das áreas urbanas.

Os telefones celulares tornaram-se praticamente indispensáveis no dia a dia das grandes cidades. As pessoas carregam seus celulares o tempo todo, fazendo com que os registros de uso da rede de telefonia móvel sejam um indicador significativo das características de mobilidade da população, trazendo a visão espaço-temporal de ocupação das diferentes áreas urbanas como mencionado em Zhang et al. (2019) e Furno et al. (2017).

As análises nessa área de aplicação baseiam-se na premissa de regularidade do comportamento de mobilidade dos seres humanos, assim os padrões espaço-temporal das atividades podem revelar os diferentes usos funcionais das áreas, por exemplo, mais atividades acontecem durante o dia no meio da semana em áreas de

escritórios, enquanto mais atividades acontecem durante a noite e finais de semana em áreas de lazer, como exemplificado em Mao et al. (2016).

Energia – Consumo de energia

Observa-se o estudo da demanda de energia elétrica como uma segunda importante área de aplicação para a análise de dados de telefonia móvel. Com base na dinâmica da população, obtida dos dados agregados de uso da rede de telefonia móvel combinada com o histórico de consumo de energia é possível, por exemplo, realizar previsões da demanda de energia elétrica diária e por hora como em Wheatman et al. (2016) e Selvarajoo et al. (2018), dos picos como em Bogomolov et al. (2016) ou obter subsídios para planejamento de infraestrutura de energia como em Martinez-Cesena et al (2015).

A eficiência energética é tida como um fator chave para a construção de sociedades sustentáveis. Com o crescimento da população, aumento da renda e industrialização, o consumo mundial de energia cresce a cada ano, sendo projetado para 2050 um aumento de cerca de 50% do consumo mundial (“U.S. Energy Information Administration (EIA),” n.d.), pressionando as concessionárias e os governos a atenderem à demanda, trazendo desafios importantes na geração, distribuição e armazenamento de eletricidade.

A previsão acurada da demanda é uma preocupação global na busca pelo aumento dos níveis de eficiência energética, trazendo um impacto significativo na prevenção de sobrecargas e no armazenamento eficiente de energia, além de fornecer informações relevantes para os processos de tomada de decisão em áreas como geração e compra de energia. Ainda, o conhecimento detalhado das necessidades de energia com uma boa resolução espaço-temporal é crucial para o planejamento adequado da infraestrutura energética de um país, em particular para eletrificação de regiões em desenvolvimento, onde novas infraestruturas precisam ser construídas para promover o crescimento socioeconômico com recursos limitados.

Os métodos tradicionais de previsão de demanda usualmente envolvem diversos conjuntos de dados, como previsões meteorológicas, estimativas de taxas de ocupação de edifícios ou históricos de consumo de energia obtidos das contas das distribuidoras. Essas fontes de dados muitas vezes estão disponíveis parcialmente, possuem limites de alcance e até erros em sua geração.

Além disso, esses conjuntos de dados não trazem a visão da dinâmica humana, isto é, não consideram a distribuição dinâmica da população no espaço ao longo do tempo, o que é particularmente relevante em áreas urbanas.

Nesse contexto, a compreensão da dinâmica da população e sua relação com a dinâmica da demanda de eletricidade é de grande relevância para melhoria da acuracidade das previsões.

Assim, para compreender a dinâmica humana em uma região, os registros de uso da rede de telefonia móvel são uma importante fonte de dados, trazendo informações sobre a distribuição da população variável no espaço e no tempo em larga escala.

A análise conjunta da dinâmica da população, derivada dos registros agregados da atividade da rede de telefonia móvel, com o histórico de consumo de eletricidade, subsidiam então modelos preditivos para previsão de demanda de energia com melhor precisão.

Segurança

A terceira área de aplicação observada é a área de Segurança, onde na subárea de gestão de multidões, os padrões de uso da rede de telefonia móvel no espaço e no tempo possibilitam detectar anomalias, associando-as a ocorrências de eventos de aglomerações de pessoas, como em Celes et al. (2019) e Zhang et al. (2019). Já na subárea de previsão de criminalidade, a visão da dinâmica espaço temporal da população derivada dos dados agregados de uso da rede móvel, é correlacionada a ocorrências de crimes, contribuindo para a previsão de nível de criminalidade nas áreas urbanas, como em Bogomolov et al. (2015).

Segundo Zhang et al. (2019), uma multidão pode ser definida como uma situação coletiva na qual os indivíduos se reúnem, com um propósito comum, em um local específico, durante um intervalo de tempo.

De acordo com Celes et al. (2019), as situações de multidões que ocorrem em uma cidade são classificadas com base em aspectos espaciais, temporais e semânticos, com por exemplo casuais, programadas e de protesto. Uma multidão casual corresponde a pessoas que estão no mesmo lugar ao mesmo tempo sem terem um vínculo comum por um curto espaço de tempo, acontecendo com frequência em grandes centros urbanos como em pontos de concentração de transporte público ou

durante o lazer em parques. Já uma multidão programada está associada a pessoas que se reúnem com propósito comum bem definido, em geral por um período mais longo de tempo, como por exemplo jogos esportivos, concertos ou exposições. Uma multidão de protesto é similar à programada, porém nesse caso as pessoas possuem uma pauta de reivindicação, como um comício ou uma marcha, estando mais suscetíveis a ocasionarem tumultos.

Com o crescimento da população nos grandes centros urbanos, há uma tendência natural para a ocorrência de aglomerações em diversas situações, sendo a gestão e controle das multidões uma preocupação constante das autoridades para evitar desordem e fornecer soluções eficientes de mobilidade, mitigando os problemas relacionados à urbanização.

Esse crescimento também acarreta maior preocupação com a criminalidade, fazendo com que os governos tenham foco na segurança do cidadão e na prevenção do crime.

Assim, nessa área de aplicação, o entendimento dos padrões das atividades e dinâmica humanas são fundamentais para subsidiar o planejamento de infraestruturas de mobilidade melhores, bem como políticas de segurança e prevenção ao crime, propiciando maior bem estar social nas cidades.

Os métodos usualmente utilizados na área de segurança para acompanhamento das atividades humanas são câmeras de vigilância, e sensores dos vários tipos instalados em vias e espaços públicos. Entretanto, esses equipamentos possuem uma abrangência limitada, e ainda, o custo para instalação e manutenção desse tipo de infraestrutura em escala urbana é muito alto.

Nesse contexto, o uso dos indivíduos como sensores surge como uma grande oportunidade de superar a dificuldade de ausência de infraestrutura de monitoramento que cubra toda a cidade. Assim, o crescente uso dos dispositivos pessoais de telefonia móvel surge como uma grande oportunidade de entendimento do comportamento humano nos ambientes urbanos.

Em particular, os registros agregados de uso da rede de telefonia móvel contribuem para trazer a visão espaço temporal da dinâmica humana, sendo de grande valia tanto para os estudos de gestão de multidões como para os de previsão de criminalidade.

Nos estudos de gestão de multidões, a análise dos registros agregados de uso da telefonia móvel possibilita o estabelecimento de padrões de atividade humana e a identificação de anomalias. Nas ocorrências de anomalias, a perspectiva temporal obtida indica a janela de tempo (início e fim) do cenário, a perspectiva espacial consiste em caracterizar como a multidão ocupa geograficamente a região da cidade, e o tema é obtido com o auxílio de fontes adicionais, como redes sociais e mapas de pontos de interesse.

Já nos estudos de previsão de criminalidade, a análise das características de mobilidade urbana, derivadas dos registros agregados e anonimizados da atividade da rede de telefonia móvel, combinada com dados demográficos possibilita modelos preditivos para classificar se locais específicos têm maior ou menor probabilidade de se tornarem pontos de alta ocorrência de crimes em um futuro próximo.

4.2.2 Tipificação da contribuição dos dados da rede de telefonia móvel

Esta seção explora as características das fontes de dados tipicamente utilizadas em cada uma das áreas de aplicação identificadas.

Como já mencionado anteriormente, quando um aparelho de telefonia móvel é usado, quer seja para realizar uma chamada de voz, enviar ou receber uma mensagem, estabelecer uma conexão para troca de dados, é gerado um registro pela operadora de telefonia móvel, denominado de CDR.

Um registro CDR contém basicamente as seguintes informações: número do telefone originador, número do telefone destino, data e hora do evento, duração, identificação da antena receptora, número sequencial único, tipo de evento (chamada de voz, mensagem (SMS), conexão de dados), e outros.

A identificação da antena traz a informação de geolocalização do evento, assim, a combinação dos componentes número do telefone, antena, data e hora do evento é particularmente útil para os estudos de mobilidade humana, pois representam um usuário no tempo e no espaço.

Os dados agregados correspondem a sumarização de conjuntos de CDRs segundo algum critério específico, onde o mais comumente utilizado é por antena, e para cada antena, por tempo, trazendo então a visão do volume de atividade daquela localidade em um dado instante de tempo.

Nos estudos de mobilidade urbana, o uso de dados agregados provenientes de CDRs pode apresentar variações em termos de tipo de eventos considerados, granularidade de tempo aplicada na sumarização dos registros, tipo de informações utilizadas na sumarização, e período dos dados analisados.

Assim, para observação das fontes de dados utilizadas foram consideradas quatro características, sendo elas o tipo de registro utilizado, tipo de informação sumarizada, granularidade temporal e período de coleta. Essas características estão sintetizadas no Quadro 5 e descritas detalhadamente a seguir.

Sobre o tipo de registro

Com o advento da tecnologia 4G e amadurecimento da Internet móvel, os aplicativos de mensageria estão cada vez mais substituindo as chamadas telefônicas, tornando-se o principal método de comunicação nas áreas urbanas. As pessoas raramente usam telefonemas para se comunicar, mas conversam frequentemente através dos aplicativos de mensagens instantâneas. Além disso, o aumento de popularidade dos aplicativos de transporte, finanças e redes sociais fazem com que a vida urbana produza o consumo de tráfego de dados móveis a todo momento. Assim, os registros de uso desse tráfego contêm informações muito ricas sobre a atividade humana, fazendo com que os CDRs referentes a conexões de dados e os registros específicos de tráfego de dados móvel sejam uma fonte valiosa de informação (Zhang et al., 2019).

A cobertura da tecnologia de telefonia móvel 4G/LTE vem aumentando com o passar do tempo, e varia de localidade para localidade, sendo maior nos países desenvolvidos, onde chega a 94% da rede ("International Telecommunication Union (ITU) Statistics," n.d.), e menor nos países menos desenvolvidos, onde a média é de apenas 45% da rede. Assim, a relevância dos registros referentes ao uso de dados móveis irá depender da localidade objeto do estudo e do período de coleta.

Nos trabalhos observados, verificou-se o uso de diversas combinações de tipos de registros, onde alguns utilizam apenas CDRs de voz (Ríos & Muñoz, 2017), outros voz e mensagens (Gundogdu et al., 2016), outros todos os tipos (Selvarajoo et al., 2018), e alguns com fontes de dados mais recentes ou focados em países onde o uso de dados móvel é alto, utilizam apenas os registros de dados móvel (Miao et al., 2018).

Na área de aplicação em Energia, todos os trabalhos observados utilizam todos os tipos de registros. Já nas áreas de Planejamento urbano e Segurança, ocorrem todas as situações, onde alguns usam apenas CDRs de voz, outros voz e mensagens, outros só de dados, e por fim, há os que usam os três tipos.

Assim, como o objetivo principal é obter a visão do nível de atividade por localização geográfica, a escolha dos tipos de registros utilizados leva em consideração principalmente a disponibilidade da fonte de dados, e os eventos que fazem mais sentido para a região em estudo, de acordo com suas características socioeconômicas, comportamento dos usuários e tecnologias existentes.

Sobre o tipo de informação

O tipo de informação corresponde ao dado sumarizado utilizado para trazer a visão do nível de atividade na localidade.

Este tipo pode ser o volume de registros de CDRs, ou a somatória de tráfego de dados nos casos em que são utilizados registros de uso de dados móvel, ou o total de usuários para as situações onde a fonte de dados contém a informação de usuários únicos.

Nos trabalhos observados, o uso do volume de registros de CDRs para trazer a visão de atividade da localidade está presente em 78% deles, sendo a forma mais popular de sumarização dos dados.

O uso da somatória de tráfego de dados ocorre principalmente em trabalhos que utilizam apenas os registros de dados móvel como fonte (Xing et al., 2018).

Nos trabalhos observados na área de aplicação em Energia, é utilizado o volume de registros de CDR na maior parte, apenas um utiliza o total de usuários. Na área de aplicação em Planejamento urbano, ocorrem as três possibilidades de sumarização. Na área de aplicação em Segurança, a sumarização mais utilizada é o volume de registros CDRs, somente um trabalho, que utiliza apenas os registros de dados móvel como fonte, considera a somatória de tráfego de dados.

Sobre a granularidade temporal

A granularidade temporal indica o intervalo de tempo utilizado na sumarização, complementado a visão espaço-temporal de atividade.

A granularidade temporal mais observada nos trabalhos é a de uma hora, trazendo a perspectiva das 24 horas do dia para os 7 dias da semana.

Via de regra, esta granularidade é suficiente e adequada para a maior parte dos estudos, a exceção de alguns trabalhos na área de Segurança e Energia que utilizaram granularidade de 10 minutos no lugar de uma hora (Bogomolov et al., 2016), com o objetivo de obter uma visão refinada da atividade humana no tempo e no espaço.

Sobre o período de coleta

O período de coleta dos dados nos trabalhos observados varia de um mês a 12 meses, com predominância dos períodos curtos, de um a dois meses.

Segundo (Wu, Qiao, & Yang, 2019), o padrão de comportamento de mobilidade humana apresenta alto grau de regularidade, se repetindo semana a semana, assim, a análise de períodos de 24 horas do dia para os 7 dias da semana, é a forma mais utilizada, variando a quantidade de semanas consideradas na avaliação. Em alguns casos (Mao et al., 2016), o período de 12 meses de registros de CDRs é sumarizado em 24/7, isto é, os dias da semana de todos os 12 meses são sumarizados em uma semana de 7 dias com 24 horas para cada dia.

Para os trabalhos na área de aplicação de Planejamento Urbano, mais de 80% deles utilizam períodos de 1 a 2 meses, que são suficientes para descoberta dos padrões de mobilidade humana durante as 24 horas do dia dos 7 dias da semana.

Na área de aplicação de consumo de energia, verifica-se que os trabalhos que têm como objetivo realizar previsões de curto prazo utilizam períodos de até 2 meses, já os que buscam previsões com objetivo de planejamento de infraestrutura, utilizam períodos mais longos, podendo chegar a 12 meses.

Na área de Segurança, nos trabalhos observados são utilizados períodos curtos, de 1 a 2 meses, períodos esses adequados para estabelecer-se o padrão de comportamento humano nas localidades ao longo dos dias da semana.

Quadro 5 - Sumário da Tipificação dos dados da rede de telefonia móvel

		Área/sub-área de aplicação			
		Planejamento Urbano	Energia	Segurança	
		Uso funcional das áreas urbanas	Consumo de energia	Gestão de multidões	Previsão de criminalidade
Tipo de Registro	Chamada de voz	■ ■ ■ ■			■
	Chamada de voz + SMS	■		■	
	Chamada de voz + SMS + Dados	■	■ ■ ■ ■	■	
	Dados	■ ■ ■ ■		■	
Tipo de Informação	Volume de Registros	■ ■ ■ ■ ■ ■ ■ ■	■ ■ ■ ■	■ ■	■
	Total de Dados Trafegados	■ ■		■	
	Volume de Registros + Tráfego	■			
Granularidade Temporal	10 minutos	■	■	■ ■	
	60 minutos	■ ■ ■ ■ ■ ■ ■ ■	■ ■ ■	■	■
Período de Coleta	de 1 a 2 Meses	■ ■ ■ ■ ■ ■ ■ ■	■ ■	■ ■	■
	de 3 a 11 Meses	■	■	■	
	12 Meses	■ ■	■		

Fonte: Elaborado pelo autor

4.2.3 Técnicas de pré-processamento e análise

Esta seção apresenta uma visão das principais etapas de pré-processamento e métodos de análise observados.

Pré-Processamento

O pré-processamento consiste na preparação dos dados, oriundos dos registros de uso da rede de telefonia móvel, para a realização das análises objeto de cada tipo de aplicação.

Na observação, verifica-se que essa etapa é similar para as três áreas de aplicação identificadas, sendo composta de quatro sub-etapas. O Quadro 6 - Sumário Pré-Processamento **Erro! Fonte de referência não encontrada.** traz a visão sumarizada da etapa de Pré-processamento, e o detalhamento é apresentado a seguir.

Quadro 6 - Sumário Pré-Processamento

Etapas	Descrição	Exemplos de Técnicas
Limpeza dos dados	Eliminação de ruídos, dados duplicados, dados incompletos e outros que possam causar distorções na análise	Técnicas usuais de limpeza de dados
Mapeamento Geográfico	Mapeamento das posições geográficas das estações base (antenas) da rede de telefonia móvel para uma representação no espaço urbano	Diagrama de Voronoi
Agregação / Sumarização	Agregação dos registros de CDRs para cada uma das antenas, por dia e por hora, no padrão 7 dias/24 horas, resumindo os volumes e/ou tráfego de dados, conforme o tipo de informação utilizada	Séries Temporais
Normalização	Normalização da agregação/sumarização obtida para possibilitar a comparação dos padrões de atividade das diferentes antenas com a mesma magnitude	Técnicas usuais de normalização Z-Score

Fonte: Elaborado pelo autor

Sub-etapa 1: Limpeza dos dados

A limpeza dos dados é importante para eliminar registros duplicados ou incompletos, inconsistências, registros identificados como ruídos e outros que possam causar distorções na análise a ser efetuada.

Essa limpeza de dados não requer técnicas especiais, sendo realizada da forma usualmente aplicada em processos de tratamento de dados. Os critérios para a limpeza são definidos caso a caso.

Exemplificando, observa-se a realização da limpeza de dados em Gundogdu et al. (2016) e Zhang et al. (2019).

Sub-etapa 2: Mapeamento geográfico

A informação de localização geográfica é obtida da identificação da antena. Em geral as fontes de dados trazem as coordenadas geográficas das antenas, latitude e longitude.

Observou-se que é comum o uso de diagramas, em especial o diagrama de Voronoi, para aproximação da área abrangida por cada antena ao espaço urbano, efetuando-se então o mapeamento das antenas aos polígonos do diagrama, como por exemplo em Mao et al. (2016) e Furno et al. (2017). Assim, cada antena é associada a um polígono.

Sub-etapa 3: Agregação / Sumarização

A análise de séries temporais é amplamente adotada nos estudos de mobilidade humana urbana nas mais variadas situações, sendo um recurso valioso para possibilitar a investigação da dinâmica espaço-temporal humana.

De modo geral, nessa sub-etapa é realizada a agregação e sumarização dos registros de CDRs em séries temporais com o fim de obter-se para cada antena: o volume e/ou tráfego a cada intervalo de hora do dia (ou intervalo de 10 minutos, conforme a granularidade utilizada). É comum o uso do padrão 7 dias da semana por 24 horas.

A informação sumarizada pode ser tanto o volume de chamadas de voz ocorridas (Mao et al., 2016), como chamadas mais mensagens trocadas mais conexões de dados (Selvarajoo et al., 2018), como apenas conexões de dados (De Nadai et al., 2016) ou quantidade de tráfego de dados (Miao et al., 2018), de acordo com o tipo de informação utilizada para representar a dinâmica da população.

O número de elementos das séries temporais irá variar em função do período de dias e intervalo de tempo considerados na análise. Por exemplo, em Zhang et al (2019) é obtida uma matriz de 4.032 elementos para cada antena, onde cada elemento da série temporal contém o volume de tráfego no correspondente dia e intervalo de 10 min, para um período de 4 semanas.

Sub-etapa 4: Normalização

Uma vez geradas as séries temporais, em geral é necessário realizar a normalização da medida de atividade para que seja possível realizar a comparação dos padrões das diferentes antenas com a mesma magnitude.

A normalização pode tanto ser realizada com técnicas simples como por exemplo em Selvarajoo et al. (2018), onde a atividade normalizada de cada elemento é definida pela atividade original do elemento dividida pelo total de atividade daquele intervalo de tempo para todas as antenas, ou com técnicas mais específicas como Z-Score utilizada em Miao et al. (2018).

Análise

Na etapa de análise realizam-se as descobertas correspondentes a cada estudo.

No geral, para as análises são utilizadas técnicas estatísticas e de aprendizado de máquina, técnicas essas variáveis de acordo com o objetivo de cada estudo e área de aplicação em mobilidade urbana. As técnicas observadas estão sumarizadas no Quadro 7 e são descritas a seguir.

Quadro 7 - Sumário Análise

Área de Aplicação		Descrição Geral do objetivo da Análise	Métodos Utilizados	Exemplos de Técnicas
Área	Subárea			
Planejamento Urbano	Uso funcional das áreas urbanas	Descoberta de padrões de atividade humana ao longo dos dias da semana (dia e horário) e associação dos padrões descobertos a tipos de uso funcional das áreas urbanas, identificando áreas residenciais, comerciais, lazer, mistas e outras	Aplicação de métodos de Aprendizado de Máquina - Modelos de Aprendizado Não Supervisionado - Clusterização sobre as séries temporais	LDA (Latent Dirichlet Allocation) NMF (Non-negative Matrix Factorization) Agglomerative Hierarchical Clustering
Energia	Consumo de energia	Previsão da demanda de consumo médio de energia elétrica nas localidades ao longo da semana (dia e horário)	Aplicação de métodos de Aprendizado de Máquina - Modelos de Aprendizado Supervisionado - Regressão sobre as séries temporais	Árvores de decisão (algoritmo Random Forest de Breiman) ARIMA (Autoregressive integrated moving average)
Segurança	Gestão de multidões	Deteção de eventos de aglomerações de pessoas através da Identificação de ocorrências de atividade humana fora do padrão esperado nas localidades ao longo dos dias da semana e horários	Aplicação de métodos de deteção de anomalias de Aprendizado de Máquina sobre as séries temporais	Seasonal Hybrid ESD (S-H-ESD) MMPP (Markov modulated Poisson process)
	Previsão de criminalidade	Previsão de índice de criminalidade nas localidades	Aplicação de métodos de Aprendizado de Máquina - Modelos de Aprendizado Supervisionado - Regressão	Árvores de decisão (algoritmo Random Forest de Breiman)

Fonte: Elaborado pelo autor

Em Planejamento Urbano, na subárea de Uso Funcional das áreas urbanas, o principal objetivo é a descoberta do tipo de uso das áreas urbanas, identificando áreas residenciais, comerciais, lazer, mistas e outras. Para tal é necessário encontrar os padrões de atividade humana ao longo dos dias da semana (dia e horário) e associar os padrões encontrados aos tipos de uso funcional, classificando assim as várias áreas urbanas.

Para explorar os dados obtidos através das séries temporais e encontrar grupos com atributos similares são utilizados métodos de aprendizado de máquina não-supervisionado, em particular, métodos de clusterização. Como exemplo, tem-se o uso dos algoritmos LDA (Latent Dirichlet Allocation) (Ríos & Muñoz, 2017), NMF (Non-negative Matrix Factorization) (Mao et al., 2016) e Agglomerative Hierarchical Clustering (Zhang et al., 2019).

Na área de Energia, subárea de Consumo de Energia Elétrica, o objetivo é efetuar a previsão da demanda de energia elétrica nas localidades ao longo da semana (dia e horário). Esse tipo de análise envolve, primeiramente, o estudo da correlação entre os dados históricos de consumo de energia e os dados referentes à dinâmica espaço-temporal humana, obtidos dos registros de uso de telefonia móvel. Na sequência, para efetuar as previsões de demanda utilizando as séries temporais geradas, são utilizados métodos de aprendizado de máquina supervisionado, em particular, métodos de regressão. Como exemplo, tem-se o uso do algoritmo ARIMA (Selvarajoo et al., 2018) , e uso de árvore de decisão baseado no algoritmo Random Forest de Breiman (Bogomolov et al., 2016).

Na área de Segurança, subárea de gestão de multidões, a análise realizada busca a detecção de eventos de aglomerações de pessoas através da identificação de ocorrências de atividade humana fora do padrão esperado nas localidades ao longo dos dias da semana e horários.

Para realizar a identificação de ocorrências fora do padrão esperado, são utilizados métodos de aprendizado de máquina de detecção de anomalias aplicados sobre as séries temporais geradas na etapa de pré-processamento. Como exemplo, tem-se o uso dos algoritmos MMPP (Markov modulated Poisson process) (Gundogdu et al., 2016) e Seasonal Hybrid ESD (S-H-ESD) (Celes et al., 2019).

Agora, para a subárea de previsão de criminalidade, o objetivo da análise é prever o índice de criminalidade nas diferentes localidades urbanas. Este tipo de análise envolve um número considerável de variáveis combinando informações provenientes de bancos de dados de criminalidade, indicadores sócio-econômicos das áreas e mobilidade urbana obtida das séries temporais resultantes dos registros de uso de telefonia móvel. Para tal são utilizados métodos estatísticos para estudo de correlação entre as variáveis e métodos de aprendizado de máquina supervisionado, em particular, métodos de regressão para previsão dos índices de criminalidade nas

áreas. Como exemplo, em Bogomolov et al. (2015), são utilizados os métodos estatísticos de correlação de Pearson, para visualização e melhor entendimento do relacionamento entre as variáveis, e coeficiente de Gini, para redução e seleção do subconjunto de variáveis a serem utilizadas no modelo de análise. Como modelo de análise dessas variáveis, foi utilizado árvore de decisão baseado no algoritmo Random Forest de Breiman para prever se uma dada área será ou não um ponto de alta criminalidade no mês seguinte.

4.3 AVALIAÇÃO DOS RESULTADOS

Esta seção discute os resultados obtidos e desafios encontrados na observação das referências do estudo de caso.

De modo geral, no estudo de caso observou-se que os resultados esperados foram atingidos utilizando como fonte de dados principal os registros provenientes do uso da rede de telefonia móvel.

Esta seção resume os pontos positivos e desafios observados em relação ao atingimento dos resultados considerando os seguintes aspectos: tipos de registros utilizados; se a privacidade dos dados é um tópico de atenção; benefícios do uso de registros agregados; se são necessários dados referentes a um longo período compreendendo vários meses para a análise; disponibilidade de fontes de dados para validação ou treinamento dos modelos; a relevância da limpeza dos dados; e pontos referentes à precisão de localização.

Tipos de registros utilizados

O comportamento humano de uso de comunicação móvel é crescente, cada vez mais os indivíduos utilizam seus dispositivos móveis, independentemente da forma de uso, assim o rastro gerado por essas comunicações representa uma valiosa fonte de dados para mapeamento da mobilidade humana nas cidades.

Observou-se que os vários tipos de registros utilizados para a análise, quer sejam referentes a chamadas de voz, envio/recebimento de SMS, ou uso de dados, possibilitaram mapear a população dinâmica de uma localidade no espaço e no tempo, sendo a base para as análises desejadas com resultados positivos.

Não obstante, verificou-se que a qualidade dos resultados obtidos pode ser variável em função do nível de atividade móvel na localidade analisada, por exemplo, na área de previsão de demanda de energia, observou-se que a correlação entre uso da telefonia móvel e o consumo de energia é maior em municípios com predominância de áreas de diferentes usos funcionais. Para municípios de baixa atividade humana ou industriais, a correlação entre o uso da telefonia móvel e o consumo de energia é baixa.

Ainda, em geral, os métodos de análise observados correspondem a modelos replicáveis que podem ser aplicados a outras cidades.

O principal desafio referente a esse ponto é a obtenção dos registros de uso da rede de telefonia móvel para a realização das análises. Esses registros pertencem às operadoras de telefonia, e cabem a elas a decisão de disponibilização das fontes de dados.

Sobre a privacidade dos dados

No estudo de caso observou-se que com o uso de registros agregados de uso da rede de telefonia móvel não há informação de identificação dos usuários, e também nenhum dado pode ser rastreado até os indivíduos. Assim, não há preocupação com privacidade de dados.

Mesmo nas situações onde são coletados registros individuais como um passo inicial, esses registros são anonimizados garantindo a privacidade das informações dos usuários.

Uso de registros agregados

No estudo de caso verificou-se que o uso de registros de dados agregados de telefonia móvel trouxe os resultados desejados nas áreas de aplicação de descoberta do uso funcional das áreas urbanas, previsão de consumo de energia, detecção de aglomeração de pessoas e previsão de índice de criminalidade, não sendo necessário dados individualizados para tal.

Além disso, o tratamento dos dados de forma agregada usualmente traz menor complexidade à etapa de pré-processamento, não demandando técnicas especiais relativas a grandes volumes de dados.

Outro ponto relevante em relação ao uso de registros agregados é a maior disponibilidade de fontes de dados com esse formato, já que estes não trazem questões referentes à privacidade de dados.

Sobre o período de tempo analisado

Conjuntos de dados compreendendo períodos curtos comumente limitam as análises, quer seja por não responderem pela sazonalidade em uma escala anual, como por exemplo períodos de dois meses para estudos na área de planejamento de consumo de energia, quer seja por poderem trazer vieses sazonais, feriados ou períodos especiais de férias, como o caso de conjunto de dados que incluem período de férias utilizados para análise de uso funcional das áreas urbanas.

À exceção de estudos voltados à previsão de demanda de consumo de energia de curto prazo, em todas as outras áreas observam-se que conjuntos de dados de curto período podem acarretar limitações na análise, sendo preferível o uso de conjuntos de dados de períodos mais longos para evitar distorções.

Na área de segurança, o uso de conjuntos de dados de períodos longos possibilita também expandir a análise cobrindo vários tipos de cenários de aglomeração de pessoas.

Fontes de dados para validação ou treinamento dos modelos

Uma das dificuldades observada com frequência no estudo de caso é a falta de dados atualizados e completos para a validação das análises efetuadas ou treinamento dos modelos de aprendizado de máquina. As fontes desses dados em geral são fontes abertas, de responsabilidade dos órgãos públicos, que não possuem recursos para mantê-las atualizadas.

Como exemplos tem-se a escassez de documentação oficial de uso funcional das áreas para verificação dos resultados das análises, a carência de dados atualizados referentes à infraestrutura e informações demográficas da população.

Na área de segurança, pode-se citar fontes de dados históricos de criminalidade agregados em base mensal, sem visibilidade diária e por horário, bem como registro incompleto de eventos nas fontes públicas, por exemplo, falta de local ou datas divergentes nas várias fontes para o mesmo evento.

A relevância da limpeza dos dados

Em algumas situações pode-se ter “ruídos” no conjunto de dados que geram distorções na análise, “ruídos” esses detectados no início do processo ou apenas em etapas avançadas, quando então é necessário retornar ao início do processo, eliminar os dados que geram distorções e refazer a análise. Dependendo do tipo de “ruído”, podem ser necessárias técnicas específicas para filtragem desses dados.

Ainda, a perda significativa de registros de dados correspondentes a certas antenas pode acabar gerando falsos positivos no modelo de análise.

É recomendável então efetuar uma análise prévia do contexto da fonte de dados para buscar identificar o quanto antes se podem haver “ruídos” ou falta de registros, e as necessidades de limpeza nos dados.

Precisão de localização

A falta de precisão de localização dos CDRs pode limitar os resultados da análise. A informação de localização usualmente é referente à posição geográfica da antena, a qual abrange uma determinada área de cobertura. Essas áreas são de tamanhos diferentes nas várias regiões de uma cidade, onde há regiões de maior densidade de antenas, trazendo maior precisão, e outras onde as antenas abrangem áreas de tamanhos maiores, diminuindo então a precisão da análise. Ainda, a distribuição de antenas é diversa nas diferentes cidades.

Assim, o impacto da falta de precisão de localização dos CDRs nos resultados é dependente da natureza do estudo efetuado. Via de regra, esse impacto será maior nas análises voltadas a segurança e energia, onde qualquer melhoria neste quesito é relevante para a análise, e menor nas voltadas ao uso funcional das áreas urbanas. Mesmo nesse último caso, a melhoria da resolução de localização dos registros de uso dos dispositivos móveis possibilitaria o desenvolvimento de estudos mais detalhados do uso funcional de áreas específicas.

4.4. DEFINIÇÃO DE UM FRAMEWORK ANALÍTICO

Como discutido nas seções anteriores, as características extraídas dos dados agregados de telefonia móvel retratam a regularidade e diversidade da dinâmica humana nas diversas áreas de uma cidade, ao longo do tempo. A análise dessas

características propicia importantes descobertas sobre a mobilidade humana urbana contribuindo para aprimoramento das áreas de planejamento urbano, energia e segurança no contexto de cidades inteligentes.

Esta seção apresenta as recomendações e orientações referentes ao processo de análise dessa importante fonte de dados, consolidando a observação do estudo de caso objeto dessa pesquisa. Essas recomendações são tratadas tanto no sentido positivo, isto é, sobre o que se recomenda adotar, como negativo referente às recomendações do que deve ser evitado, e poderão ser utilizadas para apoiar trabalhos futuros nessa área.

As recomendações são apresentadas por meio da definição de um framework analítico orientado a dados, que utiliza os dados agregados de uso da rede de telefonia móvel como fonte principal, trazendo a compressão da dinâmica humana e visando colaborar com a crescente demanda de soluções de mobilidade urbana.

O framework analítico apresentado abrange estudos de detecção do uso funcional das áreas urbanas, previsão de demanda de consumo de energia elétrica e detecção de aglomeração de pessoas.

A Figura 8 ilustra o framework definido, que é composto de quatro passos, onde o primeiro passo contempla os pontos referentes à coleta dos dados provenientes dos registros de uso da rede de telefonia móvel, o segundo aos tratamentos recomendados no passo de pré-processamento visando a preparação dos dados para o passo seguinte, o terceiro refere-se à análise dos dados gerados pelo passo 2, e finalmente o último passo sobre os resultados obtidos nas três áreas de aplicação em mobilidade urbana. As recomendações apresentadas têm como foco os primeiros dois passos, e são descritas a seguir. Para o terceiro passo, o framework apresenta uma breve orientação sobre os métodos de aprendizado de máquina usados para cada área de aplicação.

Figura 8 - Framework analítico para mobilidade urbana

Fonte: Elaborado pelo autor

Primeiro passo – Coleta dos dados

O primeiro passo corresponde à coleta dos dados. Esse passo é de grande importância, pois as decisões nele tomadas afetam todo o restante do processo analítico.

Os registros de uso da rede de telefonia móvel, denominados genericamente de CDRs, têm sido amplamente utilizados para estudar a mobilidade urbana devido às seguintes vantagens: eles fornecem uma forma de obter dados de localização e temporalidade dos usuários em grandes escalas; eles possuem grande abrangência populacional; eles podem ser obtidos para diferentes países e geografias, dada sua difusão mundial; e ainda, eles fornecem um conceito objetivo de localização, a torre de telefonia móvel.

No entanto, vale ressaltar que os CDRs podem sofrer diferentes tipos de vieses, como: a posição de um indivíduo é conhecida no nível de granularidade das torres de telefonia móvel; dependendo do tipo de registro utilizado, podem haver grandes intervalos de tempo sem registro de atividade; o comportamento humano de uso da telefonia móvel pode ser variável dependendo da geografia; os níveis de atividade podem sofrer sazonalidades; representatividade da operadora de telefonia origem dos dados na localidade analisada.

Em função do exposto, são apresentadas a seguir as recomendações para a etapa da coleta dos dados.

A primeira recomendação refere-se ao tipo de registro de CDR a ser utilizado. Como o comportamento de uso da telefonia móvel tem mudado, cada vez mais as

peessoas utilizam os aparelhos celulares para acessar a Internet, utilizar aplicativos de troca de mensagens, e outros recursos que usam a comunicação de dados, e por sua vez, falam cada vez menos, recomenda-se que sejam utilizados os registros de uso de dados ou os registros de todos os tipos de utilização, assim obtendo-se uma visão que melhor refletirá o nível de atividade da população em uma certa região. Recomenda-se evitar o uso de apenas registros correspondentes a chamadas de voz ou envio/recebimento de SMS.

A segunda recomendação é sobre o período de coleta dos dados. Para evitar vieses de períodos de férias ou feriados, recomenda-se que o período de coleta seja de dois a três meses, considerando meses típicos do ano. Dependendo da área de aplicação, deve-se evitar período curtos, como por exemplo quinze dias para detecção de uso funcional das áreas urbanas, pois os dados podem conter feriados, períodos de férias ou eventos especiais que irão gerar distorções nos resultados. Também deve-se evitar períodos muito longos, como um ano, pois aumenta-se o volume de dados e assim, a complexidade de tratamento, a menos que a visão de um ano seja necessária para o tipo de estudo.

A terceira recomendação refere-se às questões de privacidade dos dados e ao nível de agregação. Embora os dados de uso da rede de telefonia celular tragam grandes benefícios para o estudo da mobilidade humana, é importante proteger a privacidade dos indivíduos. Como apresentado no estudo de caso, para as áreas de aplicação selecionadas não há necessidade de tratamento de registros individualizados. Assim, recomenda-se a coleta de dados anonimizados e já agregados fornecidos diretamente pelas operadoras de telefonia móvel, evitando-se as questões de privacidade de dados dos usuários. Ainda, a coleta de dados já agregados irá diminuir a complexidade no tratamento dos mesmos.

A quarta recomendação está relacionada à granularidade dos dados. Na maioria dos estudos, a granularidade de 60 minutos é suficiente, não sendo necessário maior nível de precisão temporal nos dados, à exceção de aplicações de detecção de aglomeração de pessoas, onde a granularidade de 10 minutos é recomendada. Assim, deve-se evitar obtenção de dados com menor granularidade desnecessariamente, pois esta menor granularidade irá fazer com que se tenha maior volume de dados, novamente aumentando então a complexidade de tratamento.

Finalmente, a quinta recomendação para esse passo refere-se à representatividade da operadora de telefonia móvel fonte dos dados. Há países onde uma única operadora de telefonia móvel tem predominância de atuação, assim a utilização dos registros oriundos dessa operadora traz uma boa representatividade da população. Já em outros países pode ocorrer a presença de várias operadoras, como é o caso do Brasil, onde há cerca de 20 operadoras, sendo que as quatro maiores totalizam 95% do Market share (“Anatel - Panorama,” n.d.). O uso de dados provenientes de uma operadora que possua baixa representatividade na área objeto do estudo não trará informações suficientes para refletir a dinâmica humana da população da região, prejudicando assim os resultados obtidos. Recomenda-se então avaliar a representatividade da operadora de telefonia móvel na área de estudo, e caso esta seja menor que 40%, buscar combinar com dados proveniente de uma segunda operadora.

Segundo passo – Pré-processamento

O segundo passo compreende as atividades de pré-processamento dos dados com a finalidade de prepará-los para a etapa de análise.

Este passo deve ser tratado com cuidado para evitar retrabalhos nos passos seguintes.

Como ilustrado na Figura 8, o passo de Pré-processamento compreende as etapas de limpeza dos dados, mapeamento geográfico, sumarizações e agregações que se façam necessárias, normalização dos dados quando aplicável ao tipo de análise a ser efetuada, finalizando com a geração das séries temporais a serem analisadas.

A limpeza dos dados é uma etapa praticamente obrigatória, assim a primeira recomendação desse passo é relativa a ela. Pode ocorrer que nos conjuntos de dados coletados das operadoras de telefonia móvel tenham registros incompletos, registros faltantes ou duplicações. Para evitar impactos na análise, recomenda-se que sejam executados processos de limpeza de dados, usando técnicas convencionais, para eliminar registros duplicados e registros incompletos.

A etapa de agregação e sumarização tem por objetivo a geração das séries temporais que serão utilizadas no passo de análise. Para essa etapa, conforme observado no estudo de caso, recomenda-se a sumarização dos volumes totais de

registros (chamadas de voz, SMS e dados) obtidos dos dados de CDR em séries temporais no padrão 7 dias da semana por 24 horas, para cada uma das torres, sendo essa a segunda recomendação para esse passo.

A construção de séries temporais com intervalos de 60 minutos é recomendada para as áreas de detecção do uso funcional das áreas urbanas e previsão de demanda de consumo de energia elétrica, não havendo necessidade de utilizar intervalos de tempos menores. Por outro lado, não devem ser usados intervalos de tempo maiores que 60 minutos, esse uso prejudicaria a análise dos dados e impactaria a qualidade dos resultados obtidos. A organização pelos dias da semana é importante pois tem-se mudança de padrões da dinâmica humana ao longo da semana, em especial aos finais de semana.

Assim, com os dados agregados em intervalos de 60 minutos, ao longo das 24 horas do dia para os 7 dias da semana, tem-se séries temporais que trazem uma boa visão da dinâmica humana. Já para aplicações de detecção de aglomeração de pessoas, recomenda-se que as séries temporais sejam de intervalos de 10 minutos.

Ainda na etapa de agregação e sumarização, a terceira recomendação refere-se à completude dos dados. Devido à algum problema na coleta, pode acontecer de haver dados faltantes de certos períodos para algumas antenas. Na geração das séries temporais, recomenda-se que seja verificada a ocorrência de “buracos” nas séries, e caso sejam encontrados, avaliar o descarte das séries correspondentes ou até mesmo da antena inteira. Não é recomendado o uso de dados incompletos pois estes podem ocasionar distorções nas análises, gerando resultados incorretos e retrabalho.

Após a etapa de agregação e sumarização, segue-se a etapa de normalização dos dados das séries temporais. A normalização dos dados deve ser realizada para que se possa comparar os padrões de atividade das células das diferentes torres com a mesma magnitude. Assim, ela deve ser realizada em análises do uso funcional das áreas urbanas, e não deve ser realizada no caso de detecção de aglomeração de pessoas. Já na área de previsão de demanda de consumo de energia, a necessidade ou não de normalização dos dados das séries temporais deve ser avaliada caso a caso em função do tipo de metodologia empregada e objetivo da análise.

Uma vez completados as etapas citadas anteriormente, as séries temporais estão prontas para serem analisadas.

Terceiro e quarto passos – Análise e Resultado

Para estes passos são apresentadas orientações referentes a cada uma das áreas de aplicação.

A detecção do uso funcional das áreas urbanas baseia-se na análise da dinâmica humana no espaço e no tempo para identificação de padrões de uso das áreas urbanas. A dinâmica humana é medida pelo nível de atividade de telefonia móvel em cada antena ao longo do tempo, representada nas séries temporais geradas no passo 2. Assim, essas séries temporais devem ser submetidas a processo de análise que possibilite a identificação de padrões. Para esse processo são então utilizadas técnicas de clusterização de modelos não supervisionados de aprendizado de máquina. A aplicação dessas técnicas possibilita identificar os padrões de atividade humana nas diferentes regiões e então classificar cada uma delas segundo os padrões encontrados, trazendo como resultado a categorização das áreas urbanas de acordo com seu tipo de uso funcional.

Os estudos de previsão de demanda de consumo de energia elétrica consideram a existência de uma correlação entre o consumo de energia elétrica e a dinâmica humana, e utilizam essa correlação para melhoria das previsões. A dinâmica humana é medida de forma análoga à exposta para a detecção de uso funcional das áreas urbanas. Nesse caso, as séries temporais obtidas no passo 2 são utilizadas para análise preditiva em conjunto com o consumo histórico de energia, empregando-se técnicas de regressão de modelos supervisionados de aprendizado de máquina. Como resultado, essas técnicas enriquecem as previsões de demanda de consumo de energia.

A detecção de aglomeração de pessoas faz uso da análise da dinâmica humana para, ao contrário do uso funcional das áreas urbanas, identificar ocorrência de situações fora do padrão. A dinâmica humana é medida da mesma forma, e sobre as séries temporais obtidas no passo 2 devem então ser aplicadas técnicas de detecção de anomalia de aprendizado de máquina. Com a aplicação dessas técnicas são identificados datas e locais com comportamento fora do padrão em termos de aglomeração de pessoas, bem como a duração dessas ocorrências. Essas descobertas devem ser combinadas com outras fontes de dados para completo entendimento das ocorrências.

5 DISCUSSÃO

O estudo realizado mostrou o potencial da análise dos dados de uso da rede de telefonia móvel. Esses dados possuem uma abrangência sem igual para o entendimento da dinâmica humana quando comparada a outras fontes de dados.

Na pesquisa bibliográfica efetuada neste trabalho, observou-se que outras fontes também utilizadas para mapeamento da dinâmica humana são dados geolocalizados da rede social Twitter e dados de utilização de taxi. Esses dados algumas vezes são mais fáceis de ser obter, porém não possuem a mesma abrangência que os dados de uso da rede de telefonia móvel.

Trazendo em números, segundo a Organização das Nações Unidas, a população mundial é estimada em 7,7 bilhões de habitantes, e o número de assinantes de telefonia celular segundo (“International Telecommunication Union (ITU) Statistics,” n.d.) é de 8,3 bilhões de assinantes, o que corresponde a 108 assinantes por 100 habitantes, ou seja, uma taxa de penetração de 108%. Esses números vêm crescendo ano a ano, em especial nos países em desenvolvimento.

A Figura 9 ilustra o crescimento do número de assinantes de telefonia móvel e taxa de penetração, e a Figura 10 apresenta a evolução desse nos países em desenvolvimento comparada com os países desenvolvidos.

Figura 9 - Crescimento do número de assinantes de telefonia móvel e taxa de penetração

Fonte: (“International Telecommunication Union (ITU) Statistics,” n.d.)

Figura 10 - Evolução dos assinantes de telefonia móvel nos países em desenvolvimento

Fonte: ("International Telecommunication Union (ITU) Statistics," n.d.)

Em relação à rede social Twitter, em 2019, o número de usuários móveis dessa rede social no mundo era estimado em 330 milhões ("Digital 2019 Global - We Are Social," n.d.), o que corresponde a uma taxa de penetração de apenas 4,3% da população mundial. Já os dados geolocalizados de utilização de taxi estão restritos à parcela da população que utiliza esse tipo de transporte, correspondendo então a uma amostra limitada.

Assim, análises que utilizam como fonte os dados de uso de telefonia móvel para entendimento da dinâmica humana têm alcance incomparável a outras fontes de dados.

Em relação ao framework analítico orientado a dados definido neste trabalho, um ponto importante a destacar é que esse pode também ser aplicado a outras fontes de dados que trazem informações sobre a dinâmica humana no espaço e no tempo, como os dados geolocalizados do Twitter e utilização de taxi, mencionados anteriormente e outros que possuam características similares. Os passos indicados no framework serão análogos, onde os dados obtidos do Twitter, por exemplo, deverão ser processados para obtenção das séries temporais correspondentes à dinâmica humana, e então continua-se com os passos seguintes.

Ainda, no estudo realizado verificou-se que a análise dos dados de uso da rede de telefonia móvel propicia diversos tipos de aplicação além das enfatizadas nessa pesquisa. Uma área em destaque, é o controle de epidemias, onde os dados de telefonia móvel têm sido de grande valia para ações relacionadas a COVID-19,

podendo-se citar por exemplo, a gestão do índice de isolamento social no estado de São Paulo, conforme citado em (“Isolamento | Governo do Estado de São Paulo,” n.d.).

6 CONCLUSÃO

Este estudo apresentou uma visão do potencial da análise de dados agregados do uso da telefonia móvel para a observação espaço-temporal da mobilidade humana urbana, alcançando seus objetivos.

Verificou-se, neste estudo, que esta é uma área com grande atividade de pesquisa, com possibilidades de aplicação em diferentes domínios de cidades inteligentes, como planejamento urbano, transporte, saúde, segurança, infraestrutura e outros.

Foi evidenciado que os registros de uso da telefonia móvel são uma fonte de dados importante na área da pesquisa em mobilidade humana urbana, em particular, constatou-se que os conjuntos de dados agregados contribuem para análises importantes na detecção do uso funcional das áreas urbanas, no planejamento de infraestrutura de fornecimento de energia elétrica para a população urbana, no entendimento do comportamento da população nos eventos de aglomerações e na previsão de índices de criminalidade.

Foram resumidos alguns trabalhos recentes nas áreas mencionadas em tabelas de referência, que fornecem informações detalhadas sobre as fontes de dados empregadas, os métodos e técnicas utilizados na análise, e os principais desafios encontrados.

O principal resultado deste estudo foi a definição de um framework analítico orientado a dados, composto de quatro passos, iniciando na coleta dos dados referentes ao uso da telefonia móvel, seguido pelo pré-processamento para preparação dos dados para o passo seguinte de análise, e finalizando com os resultados obtidos, com foco nas áreas de detecção do uso funcional nas áreas urbanas, previsão de demanda de energia elétrica e gestão de multidões.

O framework analítico definido é acompanhado de recomendações e orientações referentes ao processo de análise dessa importante fonte de dados, recomendações essas tanto no sentido positivo como negativo. Esse framework poderá ser utilizado para apoiar trabalhos futuros nessa área, buscando colaborar com a crescente demanda de soluções de mobilidade urbana.

Verificou-se no estudo realizado que a etapa inicial de definição do tipo de fonte de dados mais apropriada para o estudo a ser realizado, e sua abrangência tanto

em termos de período de tempo como representatividade da região analisada é um ponto-chave de sucesso para a obtenção do conhecimento desejado.

O estudo realizado e o framework analítico definido tiveram foco no tratamento de dados agregados de uso da rede de telefonia móvel e às áreas de aplicação citadas, abordando com maior profundidade as características das fontes de dados e o passo de pré-processamento para preparação dos dados para análise.

Por meio da pesquisa bibliográfica realizada neste estudo e dos resultados apontados por ele, identifica-se oportunidades tanto de aprofundamento como de aumento de abrangência da temática tratada no trabalho realizado. Assim, elencam-se algumas sugestões de trabalhos futuros sobre o assunto.

A primeira sugestão como trabalho futuro é tratar a análise de CDRs individuais. A análise de CDRs individuais permitiria extrair informações adicionais como o tempo de permanência dos indivíduos nos locais, raio de locomoção e identificação de segmentos específicos da população (por exemplo, moradores vs. visitantes). Essas informações adicionais complementariam as descobertas efetuadas, em particular no campo de uso funcional das áreas urbanas e gestão de multidões, possibilitariam maior refinamento nas análises. Por outro lado, esse tipo de análise pode trazer questões de privacidade de dados que deverão ser tratadas com cautela.

Uma segunda sugestão refere-se a ampliar o estudo para a área de aplicação de transporte na temática de mobilidade urbana, onde, de acordo com a pesquisa bibliográfica realizada, os dados de uso da telefonia móvel são também de grande relevância, verificando-se então a aderência do framework definido a essa área.

Outras sugestões como trabalhos futuros remetem ao aprofundamento das técnicas de aprendizado de máquina utilizadas. Como as técnicas utilizadas são diferentes para cada uma das áreas de aplicação, sugere-se que sejam abordadas em trabalhos futuros específicos.

Por fim, uma última sugestão é a ampliação para outras áreas de aplicação da análise de mobilidade urbana em cidades inteligentes, como por exemplo a gestão de epidemias na área de saúde, dinâmica de visitação de locais na área de turismo, estudos demográficos e outras.

REFERÊNCIAS

- Agrawal, D., Das, S., & El Abbadi, A. (2011). Big data and cloud computing. *Proceedings of the 14th International Conference on Extending Database Technology - EDBT/ICDT '11*, 530. <https://doi.org/10.1145/1951365.1951432>
- Anatel - Panorama. (n.d.). Retrieved March 24, 2020, from <https://www.anatel.gov.br/paineis/acessos/panorama>
- Bergamaschi, S. (2014). Big data analysis: Trends & challenges. *2014 International Conference on High Performance Computing & Simulation (HPCS)*, 303–304. <https://doi.org/10.1109/HPCSim.2014.6903699>
- Bhatia, A., & Vaswani, G. (2013). Big data: A review. *International Journal of Engineering Sciences & Research Technology" (IJESRT) ISSN: 2277-3878*, 2(8), 2102–2106. Retrieved from http://www.ijesrt.com/issues_pdf_file/Archives_2013/august-2013/31.pdf
- Big_data* @ *en.wikipedia.org*. (n.d.). Retrieved from http://en.wikipedia.org/wiki/Big_data
- Bogomolov, A., Lepri, B., Larcher, R., Antonelli, F., Pianesi, F., & Pentland, A. (2016). Energy consumption prediction using people dynamics derived from cellular network data. *EPJ Data Science*, 5(1), 13. <https://doi.org/10.1140/epjds/s13688-016-0075-3>
- Bogomolov, A., Lepri, B., Staiano, J., Letouzé, E., Oliver, N., Pianesi, F., & Pentland, A. (2015). Moves on the Street: Classifying Crime Hotspots Using Aggregated Anonymized Data on People Dynamics. *Big Data*, 3(3), 148–158. <https://doi.org/10.1089/big.2014.0054>
- Calabrese, F., Ferrari, L., & Blondel, V. D. (2014). Urban Sensing Using Mobile Phone Network Data: A Survey of Research. *ACM Computing Surveys*, 47(2), 1–20. <https://doi.org/10.1145/2655691>
- Celes, C., Boukerche, A., & Loureiro, A. A. F. (2019). Crowd Management: A New Challenge for Urban Big Data Analytics. *IEEE Communications Magazine*, 57(4), 20–25. <https://doi.org/10.1109/MCOM.2019.1800640>
- Cuzzocrea, A., Saccà, D., & Ullman, J. D. (2013). Big data. *Proceedings of the 17th International Database Engineering & Applications Symposium on - IDEAS '13*, 198–203. <https://doi.org/10.1145/2513591.2527071>
- DATAVERSITY - Learn Data Management at DATAVERSITY. (n.d.). Retrieved May 13, 2015, from <http://www.dataversity.net/>
- De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., & Lepri, B. (2016). The Death and Life of Great Italian Cities. *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, 413–423. <https://doi.org/10.1145/2872427.2883084>

- Digital 2019 Global - We Are Social. (n.d.). Retrieved September 13, 2020, from <https://wearesocial.com/digital-2019-global>
- Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review*, 1–38. <https://doi.org/10.1093/nsr/nwt032>
- Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., & Vinué, G. (2017). Using Twitter Data to Estimate the Relationship between Short-term Mobility and Long-term Migration. *Proceedings of the 2017 ACM on Web Science Conference - WebSci '17*, 103–110. <https://doi.org/10.1145/3091478.3091496>
- Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. *Interactions*, 19(3), 50. <https://doi.org/10.1145/2168931.2168943>
- Furno, A., Faouzi, N.-E. El, Fiore, M., & Stanica, R. (2017). Fusing GPS probe and mobile phone data for enhanced land-use detection. *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 693–698. <https://doi.org/10.1109/MTITS.2017.8005601>
- Furno, A., Fiore, M., Stanica, R., Ziemlicki, C., & Smoreda, Z. (2017). A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas. *IEEE Transactions on Mobile Computing*, 16(10), 2682–2696. <https://doi.org/10.1109/TMC.2016.2637901>
- Graells-Garrido, E., Caro, D., & Parra, D. (2018). Inferring modes of transportation using mobile phone data. *EPJ Data Science*, 7(1). <https://doi.org/10.1140/epjds/s13688-018-0177-1>
- Gundogdu, D., Incel, O. D., Salah, A. A., & Lepri, B. (2016). Countrywide arrhythmia: emergency event detection using mobile phone data. *EPJ Data Science*, 5(1), 25. <https://doi.org/10.1140/epjds/s13688-016-0086-0>
- Huang, H., Cheng, Y., & Weibel, R. (2019). Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies*, 101, 297–312. <https://doi.org/10.1016/J.TRC.2019.02.008>
- International Telecommunication Union (ITU) Statistics. (n.d.). Retrieved February 25, 2020, from <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- Ishwarappa, & Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and hadoop technology. *Procedia Computer Science*, 48(C), 319–324. <https://doi.org/10.1016/j.procs.2015.04.188>
- Isolamento | Governo do Estado de São Paulo. (n.d.). Retrieved September 13, 2020, from <https://www.saopaulo.sp.gov.br/coronavirus/isolamento>
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86–94. <https://doi.org/10.1145/2611567>
- Jahani, E., Sundsøy, P., Bjelland, J., Bengtsson, L., Pentland, A. 'Sandy,' & de

- Montjoye, Y.-A. (2017). Improving official statistics in emerging markets using machine learning and mobile phone data. *EPJ Data Science*, 6(1), 3. <https://doi.org/10.1140/epjds/s13688-017-0099-3>
- Jara, A. J., Bocchi, Y., & Genoud, D. (2014). Social Internet of Things: The Potential of the Internet of Things for Defining Human Behaviours. *2014 International Conference on Intelligent Networking and Collaborative Systems*, 581–585. <https://doi.org/10.1109/INCoS.2014.113>
- Jara, A. J., Genoud, D., & Bocchi, Y. (2014). Big Data in Smart Cities: From Poisson to Human Dynamics. *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, 785–790. <https://doi.org/10.1109/WAINA.2014.165>
- Kadar, C., & Pletikosa, I. (2018). Mining large-scale human mobility data for long-term crime prediction. *EPJ Data Science*, 7(1), 26. <https://doi.org/10.1140/epjds/s13688-018-0150-z>
- Khan, Z., Anjum, A., & Kiani, S. L. (2013). Cloud Based Big Data Analytics for Smart Future Cities. *2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing*, 381–386. <https://doi.org/10.1109/UCC.2013.77>
- Lenormand, M., Picornell, M., Cantú-Ros, O. G., Louail, T., Herranz, R., Barthelemy, M., ... Ramasco, J. J. (2015). Comparing and modelling land use organization in cities. *Royal Society Open Science*, 2(12), 150449. <https://doi.org/10.1098/rsos.150449>
- Mao, H., Ahn, Y.-Y., Bhaduri, B., & Thakur, G. (2017). Improving land use inference by factorizing mobile phone call activity matrix. *Journal of Land Use Science*, 12(2–3), 138–153. <https://doi.org/10.1080/1747423X.2017.1303546>
- Mao, H., Thakur, G., & Bhaduri, B. (2016). Exploiting mobile phone data for multi-category land use classification in Africa. *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS '16*, 1–6. <https://doi.org/10.1145/3007540.3007549>
- Martinez-Cesena, E. A., Mancarella, P., Ndiaye, M., & Schläpfer, M. (2015). *Using Mobile Phone Data for Electricity Infrastructure Planning*. Retrieved from <http://arxiv.org/abs/1504.03899>
- Miao, Q., Qiao, Y., & Yang, J. (2018). Research of Urban Land Use and Regional Functions Based on Mobile Data Traffic. *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 333–338. <https://doi.org/10.1109/DSC.2018.00054>
- Naboulsi, D., Fiore, M., Ribot, S., & Stanica, R. (2016). Large-Scale Mobile Traffic Analysis: A Survey. *IEEE Communications Surveys & Tutorials*, 18(1), 124–161. <https://doi.org/10.1109/COMST.2015.2491361>
- Naphade, M., Banavar, G., Harrison, C., Paraszczak, J., & Morris, R. (2011). Smarter Cities and Their Innovation Challenges. *Computer*, 44(6), 32–39. <https://doi.org/10.1109/MC.2011.187>

- Nations, U., of Economic, D., Affairs, S., & Division, P. (2018). *World Urbanization Prospects The 2018 Revision*. Retrieved from <https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf>
- Pappalardo, L., Pedreschi, D., Smoreda, Z., & Giannotti, F. (2015). Using big data to study the link between human mobility and socio-economic development. *2015 IEEE International Conference on Big Data (Big Data)*, 871–878. <https://doi.org/10.1109/BigData.2015.7363835>
- Raza, M. U., & Xujian, Z. (2020). A Comprehensive Overview of BIG DATA Technologies: A Survey. *ACM International Conference Proceeding Series*, 23–31. <https://doi.org/10.1145/3404687.3404694>
- Ríos, S. A., & Muñoz, R. (2017). Land Use detection with cell phone data using topic models: Case Santiago, Chile. *Computers, Environment and Urban Systems*, 61, 39–48. <https://doi.org/10.1016/j.compenvurbsys.2016.08.007>
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47. <https://doi.org/10.1109/CTS.2013.6567202>
- Selvarajoo, S., Schlapfer, M., & Tan, R. (2018). Urban Electric Load Forecasting with Mobile Phone Location Data. *2018 Asian Conference on Energy, Power and Transportation Electrification (ACEPT)*, 1–5. <https://doi.org/10.1109/ACEPT.2018.8610757>
- Talia, D. (2013). Clouds for Scalable Big Data Analytics. *Computer*, 46(5), 98–101. <https://doi.org/10.1109/MC.2013.162>
- Thuillier, E., Moalic, L., Lamrous, S., & Caminada, A. (2018). Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data. *IEEE Transactions on Mobile Computing*, 17(4), 817–830. <https://doi.org/10.1109/TMC.2017.2742953>
- U.S. Energy Information Administration (EIA). (n.d.). Retrieved March 17, 2020, from <https://www.eia.gov/>
- Vilajosana, I., Llosa, J., Martinez, B., Domingo-Prieto, M., Angles, A., & Vilajosana, X. (2013). Bootstrapping smart cities through a self-sustainable model based on big data flows. *IEEE Communications Magazine*, 51(6), 128–134. <https://doi.org/10.1109/MCOM.2013.6525605>
- Wheatman, B., Noriega, A., & Pentland, A. (2016). *Electricity Demand and Population Dynamics Prediction from Mobile Phone Metadata*. https://doi.org/10.1007/978-3-319-39931-7_19
- Wu, Y., Qiao, Y., & Yang, J. (2019). Urban Functional Area Division Based on Cell Tower Classification. *IEEE Access*, 7, 171503–171514. <https://doi.org/10.1109/ACCESS.2019.2955755>
- Xing, Z., Lin, W., Qiao, Y., Zhang, H., Sun, K., & Yang, J. (2018). Inferring Land Use Type in Urban Area with Mobile Big Data. *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 1835–1840.

<https://doi.org/10.1109/CompComm.2018.8781011>

- Yin, R. K. (2003). *Case Study Research: Design and Methods*. Retrieved from https://books.google.co.id/books/about/Case_study_research.html?id=BWea_9ZGQMwC&pgis=1
- Yuan, N. J., Zheng, Y., Xie, X., Wang, Y., Zheng, K., & Xiong, H. (2015). Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 27(3), 712–725. <https://doi.org/10.1109/TKDE.2014.2345405>
- Zhang, M., Fu, H., Li, Y., & Chen, S. (2019). Understanding Urban Dynamics From Massive Mobile Traffic Data. *IEEE Transactions on Big Data*, 5(2), 266–278. <https://doi.org/10.1109/TBDDATA.2017.2778721>
- Zhao, K., Tarkoma, S., Liu, S., & Vo, H. (2016). Urban human mobility data mining: An overview. *2016 IEEE International Conference on Big Data (Big Data)*, 1911–1920. <https://doi.org/10.1109/BigData.2016.7840811>