Data Sets, Modeling, and Decision Making in Smart Cities: A Survey

MEIYI MA, SARAH M. PREUM, and MOHSIN Y. AHMED, University of Virginia, USA
WILLIAM TÄRNEBERG, Lund University, Sweden
ABDELTAWAB HENDAWI and JOHN A. STANKOVIC, University of Virginia, USA

Cities are deploying tens of thousands of sensors and actuators and developing a large array of smart services. The smart services use sophisticated models and decision-making policies supported by Cyber Physical Systems and Internet of Things technologies. The increasing number of sensors collects a large amount of city data across multiple domains. The collected data have great potential value, but has not yet been fully exploited. This survey focuses on the domains of transportation, environment, emergency and public safety, energy, and social sensing. This article carefully reviews both the data sets being collected across 14 smart cities and the state-of-the-art work in modeling and decision making methodologies. The article also points out the characteristics, challenges faced today, and those challenges that will be exacerbated in the future. Key data issues addressed include heterogeneity, interdisciplinary, integrity, completeness, real-timeliness, and interdependencies. Key decision making issues include safety and service conflicts, security, uncertainty, humans in the loop, and privacy.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Computing methodologies** → **Modeling methodologies**; • **Computer systems organization** → **Embedded and cyber-physical systems**;

Additional Key Words and Phrases: Smart city, data sets, modeling, decision making, integrating services

ACM Reference format:
Meiyi Ma, Sarah M. Preum, Mohsin Y. Ahmed, William Tärneberg, Abdeltawab Hendawi, and John A. Stankovic. 2019. Data Sets, Modeling, and Decision Making in Smart Cities: A Survey. *ACM Trans. Cyber-Phys. Syst.* 4, 2, Article 14 (November 2019), 28 pages.
https://doi.org/10.1145/3355283

1 INTRODUCTION

The populations of large cities are growing rapidly. Rapid growth presents cities with problems such as overcrowding, resource constraints, and poor public service coverage. Cities have begun addressing the growth in many ways, including employing state-of-the-art technology such as sensing, actuation, and decision making services. For many years now, sensors in cities have been

This work was supported, in part, by NSF Grant CNS-1646470.
Authors' addresses: M. Ma, S. Preum, and M. Y. Ahmed, University of Virginia, 85 Engineer's Way, Charlottesville, VA 22903; emails: {meiyi, sp9cx, mya5dm}@virginia.edu; W. Tärneberg, Lund University, 223 62 Lund, Lund, Sweden; email: william.tarneberg@eit.lth.se; A. Hendawi, University of Virginia, 85 Engineer's Way, Charlottesville, VA 22903 and University of Rhode Island, Tyler Hall, 9 Greenhouse Rd, Kingston, RI 02881; email: hendawi@uri.edu; J. A. Stankovic, University of Virginia, 85 Engineer's Way, Charlottesville, VA 22903; email: stankovic@virginia.edu.

2378-962X/2019/11-ART14 $15.00
https://doi.org/10.1145/3355283

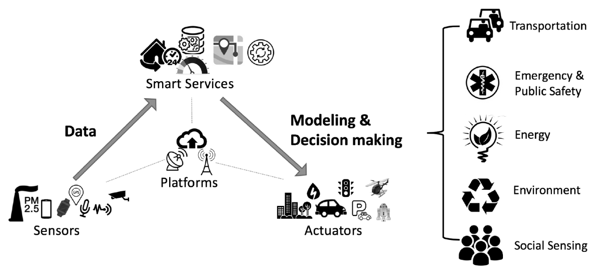

Fig. 1. Data, modeling, and decision making in smart cities.

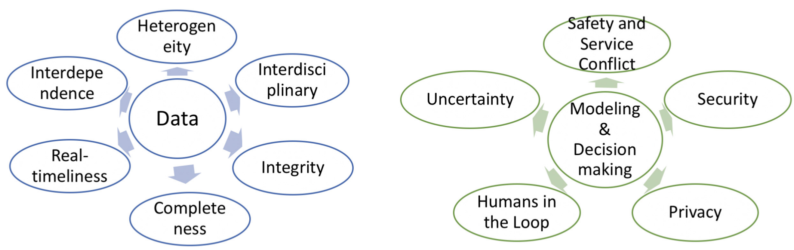

Fig. 2. Characteristics, challenges, and future work of data, modeling, and decision making in smart cities.

collecting vast amounts of data. The data have been used to construct models of the operation of the city in multiple domains and then used in various decision making tasks, which usually reflect on the actions taken on the actuators. Sensors, smart services, and actuators are running on smart city platforms, which, in a general concept, provide communications, data storage, edge/cloud computing, and so on. A general structure of smart cities is shown in Figure 1. As cities become *smarter*, increasing amounts and variety of data are being collected, more services are being provided and integrated, and more services are becoming real-time. It is projected that the economic and social impacts of smart city services will be significant [92].

There is a big gap between the data collection and data usage. On one hand, cities, especially governments, are collecting city data from different domains and encourage researchers and companies to use the data to develop smart services. On the other hand, researchers keep complaining there is not enough city data. This article is meant to fill this gap by reviewing the existing data and their characteristics, which partially account for the reason why these data are not fully exploited. We hope this article can help researchers to better understand the available data and help cities to understand the demand from research and develop better strategies to collect data.

Smart cities are projected to employ many sensors and actuators connected to the Internet. Hence, the Cyber Physical System (CPS) and Internet of Things (IoT) technologies form the basis for smart cities. Also, as more services are being deployed in cities many issues arise such as safety, privacy, security, dependability, control, manageability, and maintenance. There are some survey papers [6, 102] in these areas. Drawing from all the three areas, this survey focuses on the modeling and decision making for smart cities.

The goals for the reader (as shown in Figure 2) are to (i) develop an overall view of the scope of smart city data sets, acquire an understanding of the specifics of what is and is not contained in these current data sets, and address the characteristics, challenges, and opening future work of the data in smart cities, e.g., heterogeneity, interdisciplinary, integrity, completeness, real-timeliness, and interdependence; and (ii) expose the wealth of modeling and decision making solutions that are being applied to transportation, energy, emergency, and social sensing services. The article also highlights the methodologies and some cross-cutting issues at the forefront of modeling and decision making challenges, e.g., safety and service conflicts, security, privacy, uncertainty, and human-in-the-loop.

2 CITY DATA

2.1 Overview

A large amount of city data have been collected and published in recent years by governments, city departments, institutes, researchers, and individuals. To build a smarter city and increase accountability and responsiveness to citizens, city governments from many countries around the world are required by law to publish the city data that they collect. In New York City, it is required by Local Law 11 of 2012 that each city entity must identify and ultimately publish all of its digital public data by 2018 [22].

There already exist enormous amounts of data. For instance, there are over 183,500 data sets of American cities available on the U.S. government's open data sites with an average increase of 2,791 new data sets per month (according to data from January to June of 2016). These data sets cover a very broad range of information from individual air quality readings to overall traffic performance. To provide a more concrete understanding of the data, a set of typical examples of city data sets is shown in Table 1, which itemizes examples of the data collected from 14 representative cities in 5 domains. Some cities in Table 1, such as Åarhus, Chicago, and Paris are in the process of building a smart city, so they deploy the sensors for data collection with high density and well-chosen locations in the city. Furthermore, smart city platforms such as Arrary of Things, SmartSantander, and KM4City of Barcelona are built to obtain and display the data in real-time. These data are of great value to understand city mobility, develop new services, and improve city performance.

2.2 Existing City Data

City data are collected across a very broad set of domains, including transportation, emergency services, public safety, energy, environment, public health, social media, economics, education, telecommunication, tourism, culture, and city planning. From the varied potential purposes of city development of services, different cities emphasize different domains. For example, focusing on the development of economics and education, about 41% of the data sets from New York City [69] are related to city government and education. However, Åarhus, focuses on geographic data, environmental data, and cultural events data [1].

2.2.1 Transportation Data. Transportation data include geographical information, traffic mobility history, public transport performance, and traffic anomalies. In particular, the data come from sources such as infrastructure, traffic flow dynamics, and human-in-the-loop systems.

Infrastructure specifications, such as maps of bus stations, subway entrances, bicycle routing, and parking locations are often published as the basic information of city transportation; for instance, the transportation data from Barcelona in Table 1. Usually, they are either shown on a map visually for citizens' reference and stored in XML files for querying, analyzing, and integrating with other data sets.

Traffic flow dynamics collects traffic performance data, such as real-time traffic data, taxi trip data, and parking usage, as shown in Table 1. Traffic flow dynamics data are usually published as time-series data recordings with comprehensive traffic attributes. An example of traffic data from Åarhus is shown in Table 2, which is recorded every five minutes. Here, the sensors are deployed at two ends of one street to provide precise locations and movement data. Thus, one can recover the traffic flow dynamics from the data and correlate the traffic data with other data by aligning location and time information.

Table 1. Examples of Existing Data Sets for Cities in Different Domains

City	Transportation	Emergency & Public Safety	Energy	Environment	Social Sensing
Amsterdam [21]	Traffic, Bike share, & Accidents	Dispatches & Crime statistics		Pollution (Air & Water) & Canal water levels	Economic activity & Sentiment, Public opinion, and Tourism
Åarhus [1]	Real-time traffic data; Parking; Bicycle		Solar; Luminaires Åarhus	Cultural event; Library event; Weather; Pollution; Waste Containers	Open Data Åarhus newsletter
Barcelona [8]	Car parks; Cycling lanes; Petrol stations; Car hire; Bus stops; Car-sharing; Electric vehicle charging stations	People-involved accidents; Accidents managed by the police	Oil container recycling	Acoustic map; Weather forecast	List of media equipments and related services equipment; Shows at performing arts spaces; List of events daily
Beijing [46]	Taxi Trajectory Data set; Road networks			Air pollution: Real-time PM2.5 Air Quality Index	
Berlin [20]	Traffic & Accidents	Dispatches & Reports		Real-time pollution (Air & Water)	Economic activity & Sentiment, Public opinion
Chicago [22, 72]	Traffic tracker—Congestion estimates; Traffic counts and speed; Parking	Crimes; Snow alerts	Energy benchmarking; Energy usage(2010); Home energy score Application Programming Interface (API)	Weather; Environmental complaints	WGN-TV traffic on Twitter
Copenhagen [68]	Traffic & Bike share	Accidents, Dispatches, & Crime statistics & Outcomes	Energy prices & Public usage	Real-time pollution (Air & Water)	Economic activity, Tourism, Events, Sentiment, and Public opinion
London [23]	GB road traffic counts	London NHS A&E Performance Report; On Street Crime In Camden	Department for Transport real-time energy use	London Earth Topsoil Chemical Data	London Borough Profiles
Milan [24]	Traffic, accidents & reports			Air pollution	Economic activity & Sentiment and tourism
New York [69]	Vehicle collisions; Volume; Real-time traffic speed data; Yellow Taxi trip data;	Emergency notification; Crime data; Major felony incidents; hurricane Evacuation Centers	Energy and water data	Water quality complaints; 311 service; Air quality	*NY Times* APIs; Carpoolworld API; Yelp API; New York City (NYC) Social media usage
Paris [75]	Traffic data from permanent sensors; Taxis available referenced in mobile Taxis Paris; Parking lots	Accidents	Street lighting and traffic lights; Energy labels buildings; Volumes of water distributed	Green near me	
Santander [90]	Real-time traffic conditions; Taxi; Bus	Facility emergency (cleaning, maintenance)		Park condition; Waste; Garbage pickup; Weather; Temperature; Air quality; Noise	
Stockholm [96]	Traffic & Accidents	Dispatches & Reports	Energy prices; Demand; Central heating	Real-time pollution (Air & Water)	Economic activity & Sentiment, Public opinion
Zürich [73]	Traffic & Accidents	Dispatches & Crime statistics		Air pollution	Economic activity & Sentiment, Public opinion

Table 2. Examples of Data Set Items

Data Sets	Items
Real-time Traffic Data in Åarhus	Timestamp, Street Location, Latitude and Longitude for Start and End Point of Interests (PoIs), Counts of Vehicles, Average Speed, and Time Interval of Measurement
Criminal Data from Montgomery County	Incident ID, Dispatch Data, Class Description, Police District Name, Block Address, City, Location, Start time, End time
Austin Water Consumption	Year, Month, Postal Code, Commercial/Customer Class, Total Gallons
Energy Usage of 2010 Chicago	Community Area, Census Block, Building Type, KWH (monthly), Electricity Accounts, Therms (monthly), Gas Accounts, Occupied Units, Renter-occupied Housing Units
311 Service Request in NYC	ID, Create Data, Close Data, Agency, Complaint Type, Descriptor, Location Type, Address, Location (Latitude, Longitude)
Disaster-annotated Tweets of Italy	Tweet ID, Text, Source, Author's screen name, Author's ID, Latitude and Longitude, Time, Disaster ID, class

Table 3. Traffic Data for Taxi, Buses, Urban Trains, Bicycles and Others

City	Taxi	Bus	UrbanTrain	Bicycle	Others
Beijing [46]	Taxi Trajectory Dataset [109]	Schedule	Schedule	Public Bike [32]	
Chicago [14]		Schedule; Real-time Bus Time	Transitchicago	Divvy Bicycle Stations	Traffic Counts and Speed
London [23]	London Information System	Bus Punctuality Statistics		Cycle Parking	GB Road Traffic Counts
New York [69]	Yellow Taxi Trip Data	Schedule; Real-time Bus Time	Schedule; Subway Entries & Lines	Bicycle Routes Across New York State	Real-time Traffic Speed Data
San Francisco [70]	Travel Decision Survey	Schedule; Real-time Bus Time	Schedule	Bicycle Parking	
Seattle [71]	Taxi Transfer Dates and Values	Schedule; Real-time Bus Time	Schedule	City of Seattle Bicycle Racks	
Shenzhen [103]	Taxi Trip Data	Schedule	Schedule; Smart Card	Public Bike	

Many cities frequently publish data of different modes of transport, such as taxis, buses, urban trains, and bicycles. Some examples of available data sets are shown in Table 3. Useful information and patterns of transportation can be obtained from individual data sets as well as from integration across multiple data sets.

Human-in-the loop data, such as medallion drivers' historical archives, car- and bicycle-sharing data, and passenger waiting and transit times data are provided by many cities. Human-in-the-loop data provide data to analyze related human behaviors, and thus can lead to the improvement of transportation services.

Table 4. Examples of Different Types of Emergency Notifications for People in New York City

Type	Example
Unplanned Road Closure	"Dangerous condition due to falling glass at 200 Murray St (MN): Notification 2 sent 11/28/09 at 8:05 PM. West Street between Vesey Street and Chambers Street in Manhattan has been reopened after an earlier closure due to falling glass at 200 Murray Street. Expect residual delays."
Mass Transit Disruption	"Notification issued 4/16/14 at 2:40 PM. Due to police department activity, M trains are currently suspended in both directions between Myrtle Avenue in Brooklyn and Middle Village-Metropolitan Avenue in Queens. Consider alternate routes."
Utility	"Notification issued 1/5/14 at 12:45 AM. Con Edison is responding to a power outage in the Flatbush section of Brooklyn, including areas of ZIP code 11210, 11229, 11234."
Fire	"Notification 1 issued on 02/05/2010 at 9:00 AM. Emergency personnel are on scene of a third alarm fire at 118-39 154th St in Queens. Expect traffic delays in the area."
Environmental	"Previous 24-hour rainfall (inches) at NOAA rain gauges as of 6/02/2012 7:00 AM: JFK Airport: 1.26 LaGuardia Airport: 1 Central Park: 0.64 List of Advisories: BERGEN BASIN: CSO Advisory until 6/03/2012 3:00 AM."
Public Awareness	"Notification 1 issued 11/1/09 at 5:05 PM. There will be a 21-gun salute tomorrow 11/2/09 at approximately 8 AM from the deck of the USS New York. The ship will be in the Hudson River near the World Trade Center Site in Manhattan. Expect repetitive loud noises."
Weather	"Notification issued on 9/22/2010 at 2:30 PM. The National Weather Service forecasts the chance for strong to severe thunderstorms this evening between the hours of 5 PM and 9 PM. These storms can produce very strong and gusty winds. Please be aware of damaged trees or trees with dangling limbs."
Public Health	"Notification 1 issued on 12/4/2009 at 3:20 PM. Free H1N1 vaccine in all 5 boroughs this weekend. To check locations and see who's eligible, visit NYC.gov/flu or call 311."
School Notification	"This is a message from Notify NYC. Notification 1 issued 6/1/09 at 5:20 PM. Go to www.NYC.gov/schools or call 311 for the City Of New York's updated list of public school closures due to H1N1 flu."
Parking	"Notification 1 issued 12/19/09 at 11:58 AM. Alternate side of the street parking rules are suspended today due to the impending snow storm. For further details, please visit www.NYC.gov, or contact 311."
Missing Child / Adolescent	"Alert issued 04/30/10 at 2:25 PM. The NYS Police have issued an AMBER Alert for the abduction of a child, Shaylenn Brunson from Trinity Avenue in The Bronx at 4:30 AM. Shaylenn is a 4 year old black girl with black hair, blond streaks and brown eyes, 3 feet tall and 48 lbs."
Infrastructure	"Notification 2 issued Aug 8, 2008, 4:10 am. Because of an accident and pedestrian bridge collapse on the Bruckner Expy at Waterbury Ave, the Bruckner Expy is closed in both directions between the Cross Bronx Expy and Pelham Pkwy. Please find alternate route."

The text in the second column is extracted verbatim from [74] to demonstrate sample notifications.

2.2.2 Emergency and Public Safety Data. Emergency and public safety data include the level of safety provided to the city's populace, level of readiness of its emergency services, and occurrence of emergencies, criminal activities, accidents, and natural disasters.

Some cities reveal up-to-date individual and public emergency events obtained from emergency services. For example, the data sets of emergency notifications from New York city present timestamps, location, event type, and content of the emergency notification to the populace. Event types include severe weather, natural disasters, utility problems, mass transit disruption, and so on (see Table 4). With specific notifications, people can identify the occurrences of emergency and disruptive events and analyze the possible reasons and effects of those events from current and historical data.

Public safety data, such as criminal and accident data are another important source to monitor and to use to control city safety. Criminal data of many cities are released by the police departments. For example, Montgomery County publishes data daily on its website, providing the public with direct access to crime statistic databases of reported County crime. It reveals detailed information to the public, especially including the temporal and spatial information, items of which are shown in Table 2. Aggregated criminal data of many cities, though not as specific as event-based data, also give valuable information on the safety conditions in the city. Examples include Chicago crime data from 2001, felonies and misdemeanor crimes from the State of New York, and crime camera location data from Baltimore. Public safety data help citizens obtain immediate alerts, measure the safety in neighborhoods, distribute safety measurements and devices, and thereby build smart safety services to prevent harm to property and humans. With the development of smart real-time platforms, more timely and emergency safety data are projected to be publicly available in the near future.

2.2.3 Energy Data. Energy data focus on energy usage and production in the city, including electricity, water, gas, oil, solar energy, and other energy resources that are used by the public and individuals, as shown in the "energy" column in Table 1.

Energy consumption data sets contain the detailed electricity consumption patterns of buildings and homes (e.g., energy usage and home energy score API from Chicago, energy labels buildings from Paris) as well as the energy consumed by devices (e.g., Luminaries Åarhus). Austin Water data include commercial and residential monthly water consumption. Another important type of data is the distribution of utilities. These data are provided by some cities, such as the volume of water distributed in different areas of Paris (shown in Table 1). Example of contents of these two data sets are shown in Table 2.

Both energy distribution and consumption data sets are essential inputs of the energy models to analyze users' habit on energy usage, save wasted energy, optimize city energy distribution, and provide customized energy services.

2.2.4 Environmental Data. City environment data mainly focuses on the surroundings or conditions in which citizens live or operate. Published environmental data sets typically cover data related to air quality, weather reports, weather measurement, and green areas. Environmental data also include data associated with human-related activities, such as cultural events, waste-container status, construction areas, and citizens' voice and involvement.

Smart cities like Åarhus collect air quality data at a high density of PoIs every five minutes. Real-time data from Barcelona show the city map with distribution and availability of services including restaurants, schools, hotels, and hospitals. Mapillary [66] published a large-scale street-level image dataset containing 25 K high-resolution images from all around the world, captured at various conditions regarding weather, season, and daytime. It is annotated into 66 object categories. The data set has been broadly used to empower the machine-understanding city environment. The U.S. EPA AirNow program [3] receives real-time data from over 2 K monitoring stations, and thus provides forecast and real-time weather and air quality information for more than 300 cities in the United States, Canada, and Mexico. Water quality with total maximum daily load is collected from 622 testing sites throughout Pennsylvania to identify sources of pollution. For normal cities, forecast, current, and historical weather data and activities are provided by weather APIs such as OpenWeatherMap [76] and WunderGround [98].

The daily updated data sets of complaints in Chicago and 311 Service Requests in NYC record the voice from citizens, reflecting the complaints on noise, heating, street lights, parking, water

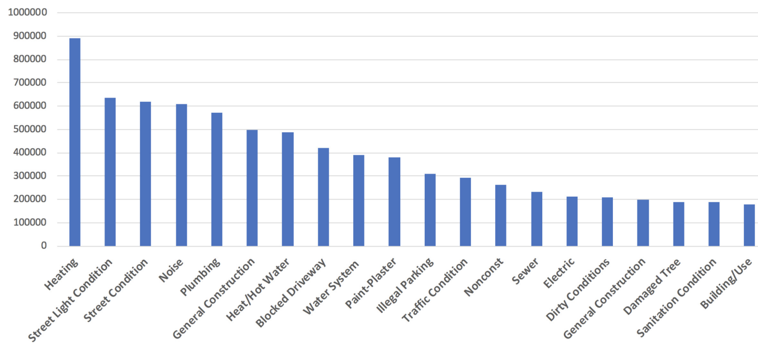

Fig. 3. Top 20 categories of NYC 311 environmental service requests (counted by the number of records).

leaks, and so on. Figure 3 indicates the top 20 categories of citizens' concerns on environment from 311 Service Requests in NYC data sets from 2010 to 2016. Hundreds of requests are added every day with very specific information, which not only includes the complaint events, but also the start and end time, locations, and detailed descriptions, as shown in Table 2. Human activity environmental data reflects the citizens' influence and degree of satisfaction on the surroundings. For example, a cultural event in downtown can cause street blockage or traffic congestion in nearby areas. Human complaints on the city environment help improve the decision making process of relevant services. Therefore, environmental data are not only meaningful for the smart environmental services, but also act as an important input source for other smart services. Meanwhile, correlating human activities with environmental data is of great significance to build human-in-the-loop smart cities.

2.2.5 Social Sensing data. Social sensing is a domain that collects data from social participants—voluntary or involuntary—and contributes that information to other domains. Typically, this type of data has traditionally been collected with surveys. Nowadays, multiple sensors (e.g., GPS, accelerometers) are integrated in mobile devices to track mobility of humans and vehicles. These devices collect data that are directly associated with human behaviors and social interactions. In smart cities, social sensing data, such as the *NY Times* and list of daily events in Table 1, give the latest news reports and event lists, while WGN-TV Traffic on Twitter updates the latest Chicago traffic accidents, delays, roadwork, transit issues, and flight info from both the public and individual sources.

Furthermore, in recent years, sensor data collection techniques and services have been integrated into social networks [2]. For example, Kosala and Adi [44] extract the real-time road traffic information in Jakarta, Indonesia, from Twitter timelines for real-time mapping; *NY Times* API provides news reports on important news and events as well as other information of all the cities in the U.S.; and based on user data from a location-based online social network in Pittsburgh, PA, Livehoods [18] collects about 18 M check-ins data. This data represents dynamic urban areas within cities. In addition, a data set [19] of 5,642 manually annotated tweets about four different natural disasters that occurred in Italy between 2009 and 2014 has been released.

By utilizing social media data, city authorities can obtain more comprehensive and timely status of a city that is difficult to gain solely from traditional sensor data and surveys. However, there are also trust and privacy issues regarding data analysis, as these data may reveal user identity.

2.3 Characteristics, Challenges, and Future Work

Comparing to other large-scale IoT systems, which are usually developed and used by one party, smart cities are built and shared by different stakeholders (e.g., governments, companies, private citizens). Sensors may be deployed first by one stakeholder and utilized by more stakeholders with more purposes than the original design. One sensor may serve multiple purposes for different services across domains. For example, a camera deployed to monitor traffic conditions at an intersection can be used by traffic optimization models, traffic violation monitors, air quality prediction models, emergency routing services, and so on in a smarter environment. This leads to the city data with very high *heterogeneity* and *interdisciplinary* uses. A smart city integrates a large number of sensors, actuators, and services across a very broad temporal and spatial range. It is very expensive (regarding both time and money) to maintain, validate, and upgrade. However, the development of technologies in smart cities evolves very fast and raises high demand for the upgrade of infrastructures and the quality of the sensors, where issues of data *integrity* and *completeness* occur. Different services using and generating new data lead to the *interdependence* among data and affect the *real-timeliness* of the services.

All of these special factors of smart cities and their influences on each other account for the complexity of smart cities, which also amplify the characteristics and escalate the challenges. In the following sections, we first discuss several key characteristics, including their importance and challenges to the development of smart cities, as well as the open questions for future work regarding how to process, integrate, and store data. In particular, we focus on seeking the reasons and solutions to shorten the gap between the collection and utilization of the data.

2.3.1 Heterogeneity. Heterogeneity of the city data refers to the data diversity in many aspects, such as,

- *Temporal and spatial granularity of collection*: Granularity is the scale or level of detail of the data in temporal and spatial domains. Most of the data streams in smart cities are highly spatial-temporal dependent, which easily leads to different sampling rates or aggregation ranges regarding time and locations. For example, the NYC open data releases the real-time traffic speed data sampled every minute by street, while the air quality data is aggregated yearly by district. This mismatch is very common in city data, and it affects the models that learn from multi-source data. For instance, it is difficult to model how the traffic influences the air quality with only yearly air quality data despite the high granularity data of traffic.
- *Application domains*: Cities have different focuses on the application domains, which leads to a variety of data sets. For example, data published by U.K. cities emphasize population information (e.g., education, age, and birth rate), Åarhus pays more attention to the city mobility (e.g., transportation), while New York City presents large amounts of city events-related data sets. Hence, it is difficult to obtain enough data across domains from one city to generalize city patterns and build standardized smart city services or platforms for multiple cities.
- *Presentation format*: Data sets are heterogeneous in the ways they are presented, such as audio, image, video or text data, raw data or processed data, and so on. Processing the data in a different format is very important for smart services. For instance, traffic services extract traffic volume and vehicle plate numbers from the video data, and autonomous vehicles learn the driving environment from street images.

Since different cities, departments, and stakeholders collect data with various protocols and intentions, heterogeneity across city data sets is unavoidable. However, heterogeneity poses significant challenges and open research directions for processing and integrating the data by smart services, which try to learn from the city data and take actions to improve city performance (see Section 3 for smarter services).

To begin with, data storage and retrieval techniques will be determined by the data types, e.g., storing and retrieving 3D models of city infrastructures involves larger amounts of resources and complexity than the coordinates of PoIs. From the perspective of information retrieval, indexing and querying heterogeneous city data depend on the application at hand. So the data storage and indexing system should efficiently support the corresponding applications. The applications might demand data representation and indexing for efficient data retrieval operation. This poses additional challenges, since, unlike traditional CPS architecture, the smart city is projected to be a platform supporting a multitude of applications that are governed and maintained by several organizations with the varying administrative capacity and information access. Thus, a smart city introduces additional technical challenges due to bureaucratic issues. Since a smart city application can support real-time large-scale data collection, appropriate data compression techniques should be adopted. For example, while collecting video surveillance data across different parts of the city for public safety applications, online stream processing and data compression should be performed to make sure the volume and velocity of the data do not surpass the data storage capacity.

Transfer learning between cities is another direction to deal with the data heterogeneity. For example, transferring the data from one city with sufficient multimodal data and labels (traffic and environment data) to another city that only has limited environmental data based on their geographic similarities [101]. People also try to infer the data to a lower granularity using its neighbor data. However, how to match features from completely different cities or locations and ensure the accuracy is still an open question.

2.3.2 Interdisciplinary. Due to the nature of the smart city research—comprehensive models built from and affecting many application domains—the research based on the city data is profoundly interdisciplinary. It usually concerns more than one branch of knowledge across different sciences and domains, including statistics, computer science, economics, political science, sociology, urban planning, transportation, environment engineering, and so on. Meanwhile, the development of IoT also promotes the interdisciplinary study of the city data. For example, Nokia Bell Labs maps the ambiance of neighborhoods in the entire city of London using social media data [85], Dantec models cycling experience in Atlanta by deploying specialized sensors to track cyclist stress, integrating noise and environmental pollution data [11].

However, the interdisciplinary study also encounters great challenges. Domain experts may not have enough knowledge of another domain. For example, environment experts may not know how to process the traffic data, which has a big influence on real-time air quality. Researchers from the public safety domain may not be good at extracting potential safety issues from social media data through natural language processing techniques. Issues brought by the interdisciplinary aspects of smart cities will escalate as the city becomes smarter in many aspects and more real deployments from the research to the field occur. For example, acoustic systems that have high accuracy to detect people's moods may have much lower accuracy in the city due to many unexpected factors, which researchers are not familiar with.

Standard policies, tools, interfaces, and platforms are potential approaches to help with the interdisciplinary issues. Companies are building smart city IoT platforms, such as, the IBM Watson IoT [100], the Azure IoT suite from Microsoft [7], and LiveLabs [9]. They integrate the data and provide a set of approaches (e.g., packages for data cleaning, text mining) to process the data. It is also possible to package the knowledge from one domain to a black box with a user-friendly interface. For example, city managers can use templates to specify city requirements in natural language without knowing anything about the formal methods used in the background for checking the requirements [55].

2.3.3 Integrity. The integrity of the city data indicates the correctness and trustworthiness of the data, especially to the city managers, developers, and citizens. The integrity of city data can be compromised during the data collection, aggregation, and processing.

- The quality of the sensors influences the accuracy of the data. Unlike a small-scale IoT application, smart cities have millions of sensors. It is very expensive and nearly impossible to maintain and validate the accuracy of all sensors all the time. For example, many parking garages deploy motion sensors to count the number of available spots, which is not accurate due to a malfunctioning of the sensors, but the managers may have no effective way to validate them in real-time.
- The environment of the sensors affects the integrity, such as the height, the depth, and the location of the *in situ* sensors. For example, the data from an air quality sensor varies when placed at different heights at the same location. Also, the data from an acoustic sensor can be very noisy if it is placed near a water pipe.
- The pre-processing process affects the integrity of the data. Before publishing, raw signal data collected from sensors are processed using algorithms from the areas of signal processing, data mining, and machine learning. As a result, the integrity of the data depends on the accuracy of these algorithms. For example, a vehicle plate recognition algorithm reads the number of a vehicle plate and records it, the integrity of which depends on the accuracy of the optical character recognition algorithm.
- Useful information is lost while data are being aggregated. In many cases, due to different reasons (such as storage demand, privacy concerns), raw data are aggregated by time intervals or locations. In the process, the information recorded at a precise timestamp and location is lost.
- Integrity may also be compromised by the security attack or deception from users on purpose. Citizens can change the GPS on their phones to protect their privacy, and criminals may attack city cameras.

City data lacking integrity diminishes the accuracy and effectiveness of smart services. Therefore, maintaining and verifying the integrity of data and understanding its provenance is important yet challenging. For instance, how to validate that smoke detection sensors are functional, and how to know if the sensors are at the right positions or facing the right directions after deployed for a long time are difficult issues? It is becoming even more challenging due to the large scale and the diversity of smart services of the smart city.

Researchers are developing strategies to ensure the integrity of the data. There are several strategies the city could apply to ensure the integrity of the data. First, adding necessary redundancy in spatial and temporal domains whenever possible, e.g., deploying multiple sensors or different type of sensors at one location with different sampling rates. People can validate the data online or offline comparing to the additional data. For example, adding cameras in the parking garage to validate the motion sensors. Second, validating the data with safety requirements using formal methods at runtime. Third, anomaly detection of online data streams might also enhance data integrity. Again, these solutions can be very time-consuming and expensive to deploy and execute at runtime.

2.3.4 Completeness. The completeness issue represents the sufficiency and the quality of the data. With the rapid development and great achievements of deep learning in CPS and IoT, an increasing number of papers apply deep learning to smart city applications to learn from the massive amounts of data available. High quality and completeness of data help the classifiers to build better models. Despite the fact that a large amount of city data has been published, its incompleteness from a service perspective exists in the following ways:

- *Incomplete data items*: It is a typical situation caused by malfunctioning of sensors or communications, and also occurs when owners of data selectively release data sets. As an example, although Åarhus collects large amounts of city data, the time span that fully overlaps across all the categories of data is only two months, which is such a small period of time to build an accurate model.
- *Incomplete attributes*: Important attributes of a type of data are not collected or revealed in the data due to various reasons. For example, in spite of being an important attribute, location information is often removed from accident and crime data sets, due to privacy concerns.
- *Incomplete contents within one domain*: As shown in Table 1, some cities collect many aspects of data within one domain, while others only collect one or even none of a particular type of data. For example, some transportation services only collect parking data from a set of locations, ignoring traffic volume data of those locations. Similarly, some services collect historical traffic volumes only on highways, ignoring traffic volumes of the rest of the streets and locations.

Incompleteness highly affects the performance of smart service models. To deal with it, on one hand, this incompleteness can be managed during data analysis before the training process. The volume and redundancy of city data can be exploited to compensate for missing data and to uncover hidden relationships. The correlation of data on the temporal and spatial domains can also be exploited to infer the missing data. On the other hand, probabilistic models are useful to reveal the uncertainties caused by the incompleteness. With probabilistic guarantees, users can be made aware of the potential uncertainties and results.

2.3.5 Real-timeliness. Concerning data collecting and processing, real-time city data streams are the data that can be accessed and processed in real-time. Only a small amount of the city data are available for real-time usage. Some real-time data APIs and dashboards are shown in Table 1. Meanwhile, cities, such as Santander and Åarhus, provide real-time data that can be processed in smart services directly. The performance and accuracy of real-time services highly depend on the quality of real-time data.

Smart services require real-time data to take real-time actions and make real-time decisions. However, most of the existing data sets go through a publication process and thus end up having a time delay from collection time to release time. Usually, the delay varies from days to years, depending on the content and privacy requirements on the data. For example, 311 service notification data are updated every day, while the latest NYC traffic data is from 2014. Without real-timeliness, there will be considerable challenges for building smart services, since most smart services take instantaneous actions responding to the city. For example, maps with navigation functions can only give an effective route based on the current real-time traffic conditions.

Realizing real-timeliness poses open questions, such as what is the required deadline and where does it come from, how to meet the deadline, how to speed up the algorithms to process a large amount of the data, how the speed and quality of the communication between services and sensors affect the real-timeliness, and so on.

2.3.6 Interdependence. The interdependence of city data indicates the relationship or interaction between different data, which commonly exists and causes significant influences on each other.

- Interdependence between data sources: the data from one data source relies on other sensors. For example, the parking service shows the occupancy of the parking garage, which is aggregated from the motion sensor on the top of each parking spot.

Table 5. The Entities Modeled Smart Cities

Domain	Entities Modeled
Transportation	Traffic flow dynamics, patterns of vehicles, pedestrians, road conditions, trajectories, EVs charging and discharging schedules, scheduling schemes, traffic speeds, smart parking, taxi dispatching, bicycle sharing, and autonomous vehicle planning.
Emergency and Public Safety	First responder strategies, emergency situations, emergency aid, nature disaster supply distribution, population mobility.
Energy	Energy demands, energy usage, renewable energy sources, building energy model, smart grids.
Environment	Monitoring, understanding, prediction, and management of city environments, primarily considering air, noise and water quality traffic emissions, health care (e.g., asthma).
Social Sensing	City events, social media messages, searching history, and so on.

- Interdependence between the data source and service data: cities not only collect data from sensors, but also from services. For example, the safety service monitors the number of people in the park, which is obtained from the safety monitor service that processes the camera data. In addition, some smart services are trained with historical data and then process the real-time data using machine learning algorithms, such as forecasting the weather or criminal facial recognition. The output data from these services highly depend on the quality of the training data.
- Interdependence between service data: some services take the output of other services as their input. For example, to predict the air quality, the environmental service unusually takes the volume and the speed of vehicles into consideration. If the traffic data generated from the traffic service are inaccurate, then so will the air quality data be.

As we have discussed for the features of heterogeneity and interdisciplinary, cities are building comprehensive models and services, which are inevitable to consider the data from multiple resources and meanwhile generate more data. This process causes interdependence.

However, the quality of the dependent data highly depends on the quality of the original data. Interdependence between services increases with the growing complexity of smart services. The verification of data accuracy and awareness of influences from other data streams are critical in the future work of smart cities.

3 MODELING AND DECISION MAKING

3.1 Overview

This section provides an overview of existing models and decision making for smart city services and systems across five different domains, including transportation, emergency and public safety, energy, environment, and social sensing.

To start with, we summarize entities modeled across these domains in Table 5, and the methodologies for modeling and decision making in papers. As shown in Table 6, integrating heterogeneous data and data analysis & mining are the most common methodologies applied. We also found that many systems use multiple methodologies to build models. In addition, we discuss the key characteristics of existing modeling and decision making systems, and the challenges and open research questions for future work.

Table 6. Methodologies for Modeling and Decision-making in Smart Cities

Methodology	Transportation	Emergency & Public Safety	Energy	Environment	Social Sensing
Dynamic programming	[4, 16]				
Quadratic programming	[16]				
Integrating heterogeneous data	[4, 31, 103]	[13, 58, 94]		[25, 38, 72]	[50, 87]
Analysis & mining	[4, 13, 15, 77, 105]	[13]	[33, 60, 63, 97, 99]		[50, 87]
Simulation	[12, 53]				
Anomalies Detection	[4]				
Physical statistical modeling	[15]		[97]	[106, 107]	
Forecasting & Scheduling	[39, 49, 51, 61, 62]				
Optimization	[62, 62, 91, 103]				
Visualization	[12]	[13]			
Crowd-sourcing	[87]				[87]
Signal Temporal Logic	[55, 56]	[55, 56]		[55, 56]	
Policies & Rules	[52]	[48]		[48, 52]	

3.2 Existing City Services Modeling and Decision Making

3.2.1 Transportation. Smart transportation systems not only promote better traffic service and citizens' daily living, but also have the potential for improving public safety and environment. State-of-the-art transportation systems are used to analyze a city's mobility, optimize the performance of traffic, and thereby provide better transportation services to build a smart city. This is done by applying models to (i) analyze and predict the traffic flow dynamics, (ii) find traffic patterns by mining the traffic trajectories, (iii) integrate heterogeneous traffic information, (iv) simulate the traffic dynamics, and (v) integrate electric vehicles (EVs) and autonomous vehicles into traditional transportation systems.

(i) As introduced in Section 2, large amounts of transportation data are collected by smart cities. Past research efforts on smart transportation use these data sets to model traffic flow dynamics, including the patterns of vehicles, pedestrians, road conditions, and city events. Comprehensive models are trying to use data from multiple sources to obtain a better image of traffic flow dynamics. For example, Anantharam et al. [4] build the Restricted Switching Linear Dynamical System to model vehicle speed and travel time dynamics. It characterizes anomalous dynamics using speed and travel time acquired of vehicles from physical sensors [70] and explains the anomalies with traffic-related incident reports from city authorities and social media data [17]. With the prevalence of deep learning techniques, researchers are starting to use deep learning models to predict traffic flow dynamics. These papers usually try to find the traffic patterns from the big traffic data considering both temporal and spatial correlations. For example, Lv et al. [51] are among the first researchers to propose a deep architecture model for traffic flow prediction using a stacked autoencoder model to extract traffic flow features and a logistic regression layer for prediction. The short-term and real-time traffic flow predictions [49] help people to make decisions on dynamic route planning.

(ii) Trajectory analysis and mining [15] are other common and fundamental ways to study traffic dynamics, including approaches such as the derivation of trajectory data, trajectory data preprocessing, trajectory data management, and so on [105]. Trace data collected from multiple sources (e.g., mobile devices, vehicles, and smart cards) are usually analyzed through clustering, classification, ranking, regression, and physical statistical modeling. Trajectories are often abstracted using data formats, such as graphs, matrices, and tensors, to which more data mining and machine learning techniques can be applied. Smart applications are built on the trajectory models to provide conveniences for citizens. For example, iDiary [31] takes GPS trajectory data generated from users' phones and turns them into textual descriptions of the trajectories to enable text-searchable activities (e.g., transportation mode, "Where did I buy books?"). To do that, iDiary compresses the semantic and clusters the trajectory of massive GPS signals in parallel to compute the critical locations of a user. Then it applies text-mining techniques on the resulting data with external knowledge to map the locations to activities. Miao et al. [61] use taxi trajectory data to develop a data-driven robust taxi dispatch framework considering spatial-temporally correlated demand uncertainties. It is evaluated with four-year taxi trip data for New York City, and the results show that it reduces the average total idle driving distance by 10.13% or about 20 M miles annually.

(iii) Strong correlations and influences exist among different transportation systems. With the understanding of traffic dynamics from heterogeneous data, more effective models and systems are built, such as smart parking [5], taxi dispatching [62], and bicycle sharing [77]. Researchers integrate and analyze heterogeneous traffic information from multiple domains of transportation to improve the performance of modeling and decision making. However, they also face many practical issues caused by the highly diverse datasets, such as heterogeneity of models, input data sparsity, or unknown ground truth. Targeting the three practical issues, UrbanCPS [103] creates three models, i.e., indirect models, sparse data models, and weighting models to infer real-time traffic speeds. Evaluating on the temporal, spatial, and contextual contexts from 42 M vehicles (e.g., buses, taxicabs, and trucks), 10 M residents and 16 M smartcards of Shenzhen, China, UrbanCPS increases the inference accuracy by 21% on average.

(iv) Simulation of urban mobility is another important and effective way to integrate, experiment with, and visualize traffic dynamics. It assists the analysis of traffic dynamics in many aspects. For example, SUMO [12], a powerful open source traffic simulation is broadly used to investigate several research topics, such as route choice, traffic optimization, and vehicular communication. With SUMO, researchers are able to (i) model inter-modal traffic systems consisting of road vehicles, public transports, and pedestrians, (ii) utilize the supporting tools, including route finding, visualization, network import, and emission calculation, and (iii) remotely control the simulation with various APIs and custom models. Ma et al. [52] uses SUMO as a platform to build a smart city simulator by implementing smart services, e.g., smart traffic services, emergency services, and environmental services, on top of SUMO. They use this simulator to predict city performance with actions from smart services and thus detect potential conflicts among them.

(v) EVs and hybrid electric vehicles are increasingly prevalent in the past two decades due to the reusable energy and low emissions to the environment [30]. The availability of EVs promotes research in modeling and decision making in smart cities, such as developing charging and discharging schedules and identifying the distribution of charging stations. With difficult challenges for handling large numbers of vehicles and their random arrivals, He et al. [39] build scheduling schemes for EVs charging and discharging using global and local optimizations to minimize the total cost of all EVs during the day and maximize the current local active set of EVs, respectively. With an increasing number of EVs, uncoordinated charging of EVs can cause power-grid problems, since it is extremely difficult to exactly predict how much power is required by household loads. Clement-Nyns et al. [16] use quadratic and dynamic programming to maximize the power grid load factor and minimize the power losses.

With the study on traffic dynamics and the development of techniques on autonomous vehicles, the internet of autonomous vehicles [35] has begun to emerge into present transportation systems. Paden et al. [78] conduct a comprehensive survey on the motion planing and control techniques for self-driving vehicles. As autonomous vehicles are programmed to make their own decisions, it is a highly distributed system comparing to traditional transportation systems. However, it also depends on the integration of information from other traffic data. The internet of autonomous vehicles [47] is designed to be intelligent with the capabilities of communications, storage, and learning to anticipate the customers' intentions with the help of the vehicular cloud. The vehicular cloud contains the models of vehicular computing, information-centric networking, and cloud resources.

3.2.2 Emergency and Public Safety. Emergency and public safety is one of the most important domains of smart city research. Cities have been pursuing ways to deal with emergencies and unsafe situations caused by natural disasters (e.g., earthquakes, hurricanes), malfunction of city facilities (e.g., power loss, bridge collapse), and human factors (e.g., car collisions, building fires) for a long time. With the prevalence of IoT, emergency and public safety systems are becoming smarter.

There are several smart modeling and decision making systems developed in the domain of emergency and public safety. In cooperation with other domains, they usually consider information of the events and behaviors of human beings, come up with strategies for potential emergency situations, build emergency and public safety services, provide first aid remotely, and make immediate decisions. For example, the IBM Intelligent operation center [13] is a system to manage emergency events including both natural and human-made disasters. This system integrates data from heterogeneous sources including the city event calendar, 911 CAD dispatch, weather services, social media, GPS and sensors, video surveillance, public works data, and traffic data. It utilizes the collected data to serve stakeholders in three stages, namely, an executive dashboard, a command center, and a first responder and field operating stage. The executive dashboard analyzes the data to alert people, detect patterns in data, and report to the command center. The command center further analyzes the data with advanced data-mining techniques for decision support. It weighs possible impacts of alternative actions/decisions and considers the parameters of solution/performance indicators (e.g., cost effectiveness, mortality, time constraints). Finally, the decisions made at the command center are forwarded to first responders and field operators for execution.

Information fusion from historical and real-time data is extremely important for decision making when it comes to emergency and public safety. Smart911 [94] signs up personal information of citizens, such as medical information, household address, and related people. If a registered person calls 911, then emergency responders are able to make decisions leveraging this information to save valuable time. Salamanderlive [89] provides a collaborative communication solution to the resolution of incidents. Users capture images (such as the wound of a patient) in real-time from the emergency scene to obtain first-aid assistance remotely.

3.2.3 Energy. City energy systems are usually dependent on national energy systems. National energy systems/services vary widely based on geographic location, availability of potential traditional and renewable energy sources, existing infrastructure and resources, economic condition of a country, and so on. Although existing works focus on different aspects of energy systems, in this survey, we only focus on the aspects that are most relevant to smart cities, such as forecasting energy demands customized to different location/countries; analyzing and predicting energy usage patterns at the individual and population levels; energy modeling for buildings [33], cities [63], and industry [60, 99]; and production and consumption of traditional and renewable energy sources. Energy issues related to transportation systems are covered in the previous section.

Energy demand forecasting models are at the heart of energy services, as they drive energy demand management and effective utilization of the energy resources. Energy demand forecasting models can be classified in various ways, such as static versus dynamic and uni-variate versus multivariate. The set of underlying techniques deployed for energy demand forecasting includes, but is not limited to, time series models, regression models, econometric models, ARIMA models, decomposition models, artificial neural network–based models, genetic algorithm–based models, and integrated models (e.g., support vector regression) [97].

Energy models for buildings are gaining increasing attention as buildings act as the unit system and contribute to the overall efficiency of the energy system of a city. An energy model refers to the simulation of energy costs, utility bills, and life cycle costs of different energy-consuming systems (e.g., air conditioning unit, lights) in an infrastructure (e.g., building, apartment complex). A comprehensive review on building energy models is found in Reference [33]. Several commercial and open source tools are available for modeling building energy; for instance, CYPETHERM, DIAL+, Elements, IDA, DesignBuilder, and eQuest [26].

Another aspect of energy systems that is transforming smart cities is improving the efficiency of energy grids. This is essential as (i) the increase in global population results in an overwhelming increase in the demand for electricity and (ii) governments across the globe are emphasizing the reduction of their carbon footprints by increasing their utilization of renewable energy sources in the power delivery chain. These complex challenges are driving the evolution of smart-grid technologies. The goal of using smart grids is to make the energy distribution and transmission more efficient, reliable, secure, meet increasing demand for energy, and reduce pollution. Specifically, smart grid technology reduces usage of fossil fuels to meet peak demand as they accommodate renewable energy sources in addition to fossil fuels. Several existing works review different aspects of smart grids. To maintain the reliability and stability of smart grid systems in the context of environmental concerns, demand side management (DSM) is used. DSM is a proactive approach where the energy consumption during peak hours is predicted, implemented, and monitored by controlling user-side consumption. Based on the scale of energy management systems, DSM applications can be categorized into three classes: residential, commercial, and industrial energy management systems [42]. Mahmood et al. focus on the technical characteristics of the communication layer of smart grids [57]. They review several communication technologies (e.g., ZigBee, Wi-Fi, Bluetooth, Z-wave, 6LoWPAN, and cellular networks) for implementing smart grids in terms of internet protocol support, power usage, data rate, range, and adoption rate. They also identify the primary challenges of implementing the communication infrastructure of smart grids, such as communication and control complexity of smart grid systems, optimizing the efficiency of different components of the integrated system, consistency, communication security, standardization, scalability, and inter-operability.

3.2.4 Environment. To build a smart environment for humans in smart cities, the present environmental models and systems are trying to maintain better monitoring, understanding, prediction, and management of city environments, primarily considering the quality of air, noise, and water.

Traditionally, environmental monitoring employs sensors and stations at a fixed and limited number of locations and are tested periodically in laboratories, which is both time- and money-consuming with a very low granularity of data. With the development of IoT and ubiquitous sensing, researchers are seeking to build cost-effective systems to obtain fine-grained environmental data in real-time. Using wireless sensor networks, Gubbi et al. [38] develop a novel scalable multi-tier noise monitoring architecture to continuously collect sound and other environmental data (e.g., temperature, humidity, and lightness) and build prototype circuitry for determining noise levels. With this platform, audio data are collected from both fixed stations and mobile devices and visualized in a phone application in real-time. The Array of Things [72] is an IoT sensing project in Chicago that deploys modular sensor boxes to collect real-time city environmental data (e.g., climate, air quality, and noise). Furthermore, Srinivas et al. [25] build a vehicular-based mobile approach to measure real-time air quality that is deployed on both public transportation and personal sensing devices. With the mobile sensing boxes and personal sensing devices, air quality is measured and uploaded to the cloud server continuously.

With big data collected from environmental monitors, approaches are developed to obtain a better understanding and thereby a more accurate prediction of city environmental variables. Despite real-time monitoring, the environmental data collected are still "past" data with limited capability to generate precaution actions. Analyzing the valuable monitoring data helps to understand the patterns and correlations between environmental variables, leading to more accurate predictions. For example, by integrating air-quality data (historical and real-time) with other related city state data (e.g., meteorology, traffic flow, human mobility, road networks, and POIs), U-Air [106] infers the real-time and fine-grained air quality (PM 2.5) data using semi-supervised learning approaches with two classifiers, i.e., a spatial classifier and a temporal classifier. The spatial classifier uses an ANN and takes space-related features such as the density of POIs and length of highways as input, while the temporal classifier uses a linear-chain conditional random field and takes time-related features in a location such as traffic and meteorology. An extended work [107] focuses on forecasting the air quality with a linear regression–based temporal predictor, a neural network–based spatial predictor, a dynamic aggregator, and an inflection predictor. These predictors are applied to model the local factors of air quality and various global factors, combining the predictions of the spatial and temporal predictors, and capturing sudden changes in air quality, respectively. The system has been deployed to provide 48-hour fine-grained hourly air quality forecasts for four major Chinese cities. One independent prediction model is trained for every six hours. The system has been deployed by the Chinese Ministry of Environmental Protection and obtains a precision greater than 0.75 in the first six hours.

Building on environmental understanding and prediction capabilities, environmental management and associated applications aim to (i) provide environment-related health advice for citizens, (ii) assist other services for decision making, and, most importantly, (iii) give real-time control over the city environments. However, only a limited number of works has appeared in this area. Some applications on smartphones offer user advice based on the environmental information, such as weather and air quality. For example, AsthmaGuide [83] provides guidance and alarms based on the real-time environment and patient's data to help asthma patients avoid exposure to high pollen counts. To assist decision making, environment modeling also helps with traffic emission estimation and anomalous events detection.

3.2.5 Social Sensing. With the contributions of the IoT and human participation in smart cities, sensing as a service and human sensing (or social sensing) have become another trend for developing services.

Decision making in smart cities is starting to take the data gathered from social media into account. With the advent of digitization, crowd-sourcing, and civic technology, various aspects of city life are influenced by social sensing. Social media acts as a passive actuator that can influence individuals, groups, or even an entire community. For example, several cities across the world have active social media presences, and they influence the decisions citizens make every day. As of

2013, in the U.S., 772 Twitter handles are run by state and local law-enforcement departments [80]. These Twitter accounts have millions of followers. The law-enforcement departments often spread notifications about important events such as, crime/incident (e.g., robbery, shooting, fire, accident), natural calamities (e.g., flash flood, earthquake, storm, snow), planned events (e.g., concert, sports match), traffic congestion, road construction, and so on. In addition, they often post safety tips to avoid unwanted events. This information supports citizens' decision making from preparation of natural disasters, mode of transportation, traffic route selection, to daily scheduling.

Though crowd-sourcing has been researched for years, using crowd-sourcing models in smart cities is still an emerging area. IBM [87] develops a crowd-sensing system for the smart cities domain and gives a case study on how it helps public safety to identify events. To start with, crowd report data is delivered to the system from various sources, such as social streams (e.g., Twitter), mobile phones (e.g., SMS), transcripts (e.g., recordings of civilian calls to the city's control center), and so on. After pre-processing, the data passes to filters such as language filters, topic filters, and context filters, then are clustered based on their semantic and contextual similarity. With weighting the event's textual features and cluster labeling techniques, a human readable summary of that event is extracted. Furthermore, the analytic component analyzes the data to extract key performance indicators and trends. At last, other physical sensor readings are integrated to improve the accuracy of the system. Liu et al. [50] build a crowd-sourcing system, ParkScan, which leverages the learned parking decision model in collaboration with the hidden Markov model to estimate background parking spot availability. ParkScan reduces by over 12.9% the availability estimation errors for all the spots during peak parking hours.

3.3 Characteristics, Challenges, and Future Work

The challenges of building models for smart cities are not only caused by the immense scale in the temporal and spatial domains and the complex dependencies among services and sensors, but also arise from other important factors. For example, a smart city is a system of systems integrating services that are often independently developed. Without knowing the context of other services, *safety and service conflicts* happen, especially at runtime. By integrating heterogeneous systems that have very different objectives, protocols, computing, and storage approaches into one system, cities are vulnerable to *security* attacks at different layers, such as the sensing layer, the computing layer, and the cloud layer. In addition, this system of systems faces strong *uncertainties* from both the external environment and between the systems supporting different domains.

Furthermore, *humans* play important roles in the city system of systems, including the source of data (sensors), sometimes acting as actuators involved in the services, and also as decision makers. On one hand, human beings are the beneficiary of the system. On the other hand, they also need to be protected from the system, such as maintaining their *privacy*, and their safety and health from *safety and service conflicts*.

Overall, from both perspectives of the system of systems and humans-in-the-loop, new challenges and open questions arise to build and deploy city models.

3.3.1 Safety and Service Conflicts. With the emergence of smart cities, governments, companies, and researchers are deploying tens of thousands of sensors in cities and developing smart services on top of them to improve city performance. However, many smart services are developed independently by different stakeholders under their own safety requirements; thus, they are unaware of actions, effects, or safety rules of other services. Serious safety issues are raised when many independently developed services are integrated into the same smart city platform [54] because of conflicts that may arise between the services.

Conflicts exist among services within or across domains. For example, the smart traffic service reduces traffic congestion and optimizes city traffic by adjusting traffic signals according to traffic conditions. At the same time, a smart event service may send a signal to the traffic lights to block a street. These two services have a conflict on a specific device, the traffic signal [53]. Ma et al. [52] present a thorough taxonomy of conflicts, categorized by device conflicts, environmental conflicts, and human conflicts.

The effects of conflicts may deteriorate performance and violate safety requirements in direct or indirect ways. For example, when managing an emergency event, an operation center [13] may cause traffic congestion nearby, violating a performance goal. In another example, a taxi dispatch system [62] may increase the air pollution in a public place (e.g., school, shopping mall) while dispatching more taxis there to provide better service. This may harm the health of air pollution sensitive people, e.g., people with asthma, thereby causing an indirect safety violation.

There are some existing works on detecting conflicts across multiple services or applications. CityGuard [52] focuses on detecting conflicts in the interventions from multiple smart city services by representing actions and effects of actions of different services using a structured format. Some runtime verification of smart cities [55] uses Signal Temporal Logic (STL) to monitor and predict traces of city states and verifies the safety and performance requirements. An interesting research direction is to detect conflicts in a dynamic complex smart city environment by interpreting and combining heterogeneous, unstructured textual, numerical, audio, and video data generated from different city services. CityResolver [56] provides a decision making support system by showing the trade-offs between different options with their violation degree to users when they are facing a conflict between actions in smart cities. However, due to the complex and dynamic actions and environment of a smart city, some conflicts are inevitable that may compromise the safety and performance of a city. Hence, run time safety monitoring to detect and resolve conflicts is also extremely important for smart cities.

There are also great challenges to address the safety and service conflicts at the early stage of the development of smart cities. To start with, there is a very limited number of safety requirements for smart cities. Even city managers do not know the proper logistics and parameters for safety. Second, it is very difficult to predict city future performance with requested actions from smart services, considering the complexity in spatial and temporal dimensions and the uncertainties in the city. Third, due to the diversity of services from different stakeholders and a large number of safety requirements from different domains, it is very difficult or even impossible to build everything into one optimization function. Fourth, all existing work applies a centralized solution, which may be problematic in a real deployment without considering different levels of detail and geographical regions. A hybrid centralized and decentralized solution should be considered in the future work. Nevertheless, we believe that developing safe and conflict-free smart city systems should be a major direction for smart cities in the future.

3.3.2 Security. Security is a crucial and challenging characteristic, since services not only collect data from the city, but also take actions on the city facilities and influence citizens' lives [104]. There are tens of thousands of things connected in smart cities, each of which is possible to compromise and perhaps jeopardize the whole system. Potential threats come from different layers of the system and can lead to serious consequences.

Sensing and Actuation Layer: The potential security issues from the sensing and actuation layers come from the nodes for data acquisition from the environment, devices for acting, and communication between them. IoT devices are developed by different stakeholders for various purposes, some of them having very little or no concern for security at all. For example, attackers compromise the high wattage IoT devices to cause local power outages and even large-scale blackouts [95]. Social media platforms usually apply sentiment analysis algorithms to the online news and posters in order to analyze social opinions (e.g., positive, negative, happy, sad) on the news or block malicious posts. However, adversary samples are injected to the posters to influence the analyzing results [34]. Social media monitoring services are not able to obtain the correct social emotions and may leave out a big event.

In addition, once sensors are deployed in the city, it is not easy to update or change the devices. However, new attacks emerge rapidly. How to maintain the security of "old" devices and develop adaptive defense strategies are still open questions.

Edge and Fog Computing Layer: Due to the overhead and delays involved with cloud computing, edge and fog computing are becoming more popular in smart city systems in recent years. Different from the cloud, edge and fog layers are composed of mobile devices mostly with limited capabilities and are more vulnerable to malicious attacks. Roman gives a comprehensive survey [88] on the security threats, challenges, and mechanisms of edge and fog computing. The attacks usually happen in the network infrastructure, edge data center, core service infrastructures, virtualization infrastructure, and user devices.

Building upon the traditional security technology, new approaches are being developed. For example, blockchain is a decentralized and distributed platform accessible by a peer-to-peer network. IoT-oriented smart city subsystems can automatically access required data from IoT devices by smart contracts in the blockchain network (e.g., Block-VN [93]), thus improving the reliability, privacy, and transparency of smart city systems. However, it is an open research question as to whether blockchain can be used at the edge and fog layers in practice due to its execution time costs.

Cloud Layer: With the prevalence and convenience of cloud computing, most of the important city data are stored and processed in the cloud. Attackers can compromise the cloud and obtain important privacy-sensitive data, sabotage the decision making algorithms, inject adversary samples to the data to influence the decision, send actions to the actuators to cause chaos in the city, and so on.

Many works have been done regarding the security of cloud computing [108]. Instead of introducing the techniques, we emphasize the challenges on cloud security raised by the characteristics of smart cities. First, data are uploaded from and used by a large number of devices. Attackers can easily get access to the cloud by disguising themselves as a device or obtain the password by hacking a vulnerable device. In addition, traditional techniques to detect an attack may not work well due to all the characteristics of city data (see Section 2.3). Furthermore, privacy-sensitive data and the variety of roles used to access the data increase the authentication challenges of the cloud security.

In summary, a smart city is a vulnerable IoT system due to its large geographical distribution, the large number of devices, and the complex communications among them. Therefore, how to ensure the security of devices is crucial for future research. To be noted, these devices (1) are highly heterogeneous regarding the operating system, RAM, functions, protocols, and deployment time and may not have the capability to install large security-based software, and (2) could be difficult to be updated once deployed and, therefore, vulnerable to new attacks.

3.3.3 Uncertainty. Smart cities exhibit a high degree of uncertainty. Services operate in open and highly dynamic environments. In this setting there are many sources of uncertainty. The environment itself is non-deterministic; trains are late, buses break down, air pollution varies by street, and so on. Human behavior is uncertain and affects services. Collected data have extraneous, missing, or incorrect values. Sensors may fail. Disruptive events occur: gas leaks, heavy rain, accidents, and so on. The actions of humans are often unpredictable. Consequently, uncertainty stands as a core challenge for effective operation of smart cities [10, 40].

In addition to the uncertainty arising from what smart cities currently have as smart components, there is also uncertainty about what they could have in the near future [10]. A smart city is not static; rather, it is a continuous process of change and improvements [43]. For example, "how smoothly will the next technology (e.g., smarter sensors) integrate with the existing sensors?" is a legitimate question of uncertainty. Thus, as uncertainty increases, the modeling and decision making in smart cities become more difficult, because decisions are based on stochastic and incomplete knowledge. In turn, confidence in the correctness and effectiveness in the models and decisions is imprecise. As a result, the quality of services in a smart city as a whole would be degraded with the improper consideration of uncertainty.

From the above, it is clear that uncertainty is a vital challenge in smart cities. Most works have addressed this challenge for a single service and often with restricted types of uncertainties. For example, Reference [82] addresses just one of type of uncertainty—the uncertainty of users' location. The authors attempt to find an efficient way to model and index imprecise positioning (i.e., inaccurate GPS or user's cloaked location) of a moving user while achieving high-quality processing for his service requests. To the best of our knowledge, there is still no research work that handles uncertainty on the global level and across services of a smart city. Thus, it remains an open research problem.

3.3.4 Humans in the Loop. Smart city services exhibit many variations of human-in-the-loop feedback control systems where the degree of human involvement varies according to system design, functionality, and implementation. In the context of smart cities, humans interact at different levels in different roles. Humans (usually government officials) make decisions about policies that are implemented for citizens. From the IoT system perspective, humans often act as sensors (e.g., social sensing, citizen governance) [86], actuators (e.g., constructing roads, rescuing people in emergency situations), and controllers (e.g., making decisions). Munir et al. identify a set of generic cyber-physical system challenges for human-in-the-loop control applications, such as handling real-time responses, deriving models of human behavior, determining ways to formally incorporate human behavior models into feedback control, and so on. [64]. There are several additional human-in-the-loop challenges in the context of smart cities as described below.

Humans as actuators: Although in existing city services automated actuation is limited (e.g., controlling traffic, parking), in the future it is projected to increase and replace humans in many cases (e.g., package-delivering drones, self-driving cars). One potential challenge here is to make the interactions between humans and autonomous technology safe and conflict-free. There are additional challenges when humans act as actuators, e.g., predicting certain human behaviors or actions at the individual and community levels. For example, it is necessary to predict travel patterns of passengers across different days of the week to adapt to dynamic demand and increase efficiency of transport systems.

Humans as sensors: While social sensing offers robust, ubiquitous, and intelligent sensing [2], this may also introduce intentional (e.g., spreading a rumor or fake news in social media) and unintentional errors. Humans are often monitored to collect data, design services, or operate services. While there are a lot of existing solutions to monitor individual behavior, it is still challenging to detect collective human behavior (e.g., activity detection in a multi-person home or in a crowded train station). Another open question regarding sensing humans is how to infer knowledge from the sensed data. Current technology focuses more on collecting a wide variety of data from humans (e.g., social media activity, activities of daily living, physiological data), but often lacks accurate inference techniques. Inference is essential to understand human behavior and use the knowledge in feedback loops.

Humans as decision makers: Decision making in the context of a smart city introduces several challenges. First, collaborative decision making process can result in conflicting opinions about a decision. As different experts with different backgrounds/motivations decide differently, conflict resolution in such cases is challenging. Another challenge is subjectivity and bias of the decision makers [41].

Second, deciding the metrics of outcome of a decision is another challenge. There may be multiple metrics involved (e.g, saving resources, causing less harm to environment) and there might be conflicting cases where optimizing one metric can negatively impact another metric [65]. This may also involve moral and ethical dilemmas regarding different policies [37], e.g., decision making logic for autonomous vehicles to minimize casualty and resources.

Finally, one of the most difficult challenges of smart cities is predicting the effects of a decision or action. For example, while planning for a rescue in an emergency event in a city, there might be multiple options to pursue. To decide optimally, one needs to analyze the effects of each of the potential options and select the one that may optimize the performance metric(s). Effects can be primary, secondary, or tertiary [52, 81]. Determining effects require accurate and timely modeling of the system, which might be difficult. When determining the effects of a decision it is also essential to detect the severity of an effect, as it can aid conflict resolution. In addition, decision making often requires predicting the time horizon of effects as well. In addition, for decision making in smart cities, it is often essential to consider not only the effects on individuals, but also the effects on a group of people or community. While existing human-centric applications heavily focus on personalization [67], future smart city applications will require consideration of both individual and collective outcomes.

Thus, humans, as decision makers, need to weigh the effects of a decision from different aspects (e.g., individual effect vs collective effect, degree of severity of an effect, the time horizon of the effect) while minimizing human errors and biases.

3.3.5 Privacy. One of the biggest challenges in smart cities is protecting individuals' privacy [28, 29]. This challenge becomes more difficult as people become more connected to technology elements, e.g., sensors, and use more smart services such as location-aware services. To receive any type of service in a smart city, the user has to release part of his private information. As an example, for a smart transportation service, to determine the nearest bus station, a rider's current location must be sent to the service provider. In smart health, to obtain advice from an on-line expert physician, a patient needs to share his medical records and perhaps current physiological information collected from wearable devices over the Internet.

To protect privacy yet permit use of services, researchers introduce a number of approaches to trade off the level of privacy with the value of services, such as cryptography, perturbation, anonymization, and cloaking: cryptography [27, 36] that turns data elements into unreadable format to anyone except those who have the decryption key; perturbation that substitutes the actual data with synthetically generated data. This approach can be seen as adding noisy readings to the original data [45, 84]. Anonymization [79] aims at hiding the user's identity when sharing his information to request a service. This can be performed by partial removal of the identifier data element that causes a user to be unrecognizable among a number of other users. For example, hide the Social Security number except the last four digits. Cloaking [59] is mainly used in location-aware services, e.g., finding the nearest gas station. Cloaking defines a user's location in a larger area, e.g., zip code, rather than the exact position, e.g., latitude and longitude.

While there is ongoing research on privacy preservation in general, there is still a lack of handling privacy holistically over all services within the context of smart cities. Privacy policy languages are needed that are understandable and easily set by users, and implementation support is also needed that dynamically enforces the policies. It is inevitable that conflicts will arise across policies, so techniques to detect and resolve policy conflicts are also required.

4 CONCLUSION

The main goal of the article is to give the reader a fairly comprehensive view of state-of-the-art in smart city data, modeling, and decision making. This article presents a survey of the data sets being collected across multiple domains by 14 smart cities. It also discusses the characteristics, associated challenges, and open research directions of the city data. The article also discusses modeling and decision making for smart city services with emphasis on their modeling entities, methodologies, capabilities, and limitations. It highlights five overarching challenges (e.g., safety and service conflict, security, privacy, human-in-the-loop, and uncertainty) and the future research questions raised by them.

REFERENCES

[1] Smart Aahurs. 2017. ODAA: Open Data Aarhus. Retrieved from: https://www.odaa.dk/dataset.
[2] Charu C. Aggarwal and Tarek Abdelzaher. 2013. Social sensing. In *Managing and Mining Sensor Data*. Springer, 237–297.
[3] Airnow. 2017. AirNow. Retrieved from https://airnow.gov/.
[4] Pramod Anantharam, Krishnaprasad Thirunarayan, Surendra Marupudi, Amit Sheth, and Tanvi Banerjee. 2016. Understanding city traffic dynamics utilizing sensor and textual observations. In *Thirtieth AAAI Conference on Artificial Intelligence*.
[5] Yacine Atif, Jianguo Ding, and Manfred A. Jeusfeld. 2016. Internet of things approach to cloud-based smart car parking. *Procedia Comput. Sci.* 98 (2016), 193–198.
[6] Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010. The internet of things: A survey. *Comput. Netw.* 54, 15 (2010), 2787–2805.
[7] Microsoft Azure. 2016. Azure IoT Hub Documentation. Retrieved from: https://azure.microsoft.com/en-us/documentation/services/iot-hub/.
[8] Tuba Bakıcı, Esteve Almirall, and Jonathan Wareham. 2013. A smart city initiative: The case of Barcelona. *J. Knowl. Econ.* 4, 2 (2013), 135–148.
[9] Rajesh Balan, Archan Misra, and Youngki Lee. 2014. Livelabs: Building an in-situ real-time mobile experimentation testbed. In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*. ACM, 14.
[10] Michael Batty, Kay W. Axhausen, Fosca Giannotti, Alexei Pozdnoukhov, Armando Bazzani, Monica Wachowicz, Georgios Ouzounis, and Yuval Portugali. 2012. Smart cities of the future. *Euro. Phys. J. Spec. Top.* 214, 1 (2012), 481–518.
[11] T. J. Becker. 2017. Interdisciplinary approaches to urban challenges are creating smart cities. *Res. Hor. Mag.* (June 2017).
[12] Michael Behrisch, Laura Bieker, Jakob Erdmann, and Daniel Krajzewicz. 2011. SUMO–simulation of urban mobility: An overview. In *Proceedings of the Third International Conference on Advances in System Simulation (SIMUL'11)*. ThinkMind.
[13] Arundhati Bhowmick, Eduardo Francellino, Luis Glehn, Robert Loredo, Pam Nesbitt, Shi Wei Yu, and Sunil Mishr. 2012. IBM intelligent operations center for smarter cities administration guide. IBM Redbooks.
[14] Chicago. 2019. Chicago Data Port. Retrieved from https://data.cityofchicago.org/.
[15] Gang Pan, Guande Qi, Wangsheng Zhang, Shijian Li, Zhaohui Wu, and Laurence Tianruo Yang. 2013. Trace analysis and mining for smart cities: issues, methods, and applications. *IEEE Communications Magazine* 51, 6 (2013), 120–126.
[16] Kristien Clement-Nyns, Edwin Haesen, and Johan Driesen. 2010. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *IEEE Trans. Power Syst.* 25, 1 (2010), 371–380.
[17] Metropolitan Transportation Commission. 2019. 511 SF Bay. Retrieved from: 511.org.
[18] Justin Cranshaw, Raz Schwartz, Jason I. Hong, and Norman Sadeh. 2012. The Livehoods Project: Utilizing social media to understand the dynamics of a city. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*. 58.
[19] Stefano Cresci, Maurizio Tesconi, Andrea Cimino, and Felice Dell'Orletta. 2015. A linguistically driven approach to cross-event damage assessment of natural disasters from social media messages. In *Proceedings of the 24th International Conference Companion on World Wide Web*. ACM.
[20] Berlin Open Data. 2019. Berlin Open Data. Retrieved from: https://daten.berlin.de.
[21] Data.amsterdam.nl. 2019. Gameente Amsterdam. Retrieved from: http://data.amsterdam.nl.

[22] Data.gov. 2017. The Home of the U.S. Government's Open Data. Retrieved from: https://www.data.gov/.
[23] data.gov.uk. 2017. Opening up Government in UK. Retrieved from: https://data.gov.uk/.
[24] Dati.comune.milano.it. 2019. Dati Comune Milano. Retrieved from: http://dati.comune.milano.it.
[25] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. 2013. Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. ACM, 15.
[26] Best Directory. 2017. Building Energy Software Tools. Retrieved from: http://www.buildingenergysoftwaretools.com.
[27] Amit Dua, Neeraj Kumar, Mukesh Singh, M. S. Obaidat, and Kuei-Fang Hsiao. 2016. Secure message communication among vehicles using elliptic curve cryptography in smart cities. In *Proceedings of the International Conference on Computer, Information and Telecommunication Systems (CITS'16)*. IEEE, 1–6.
[28] Lilian Edwards. 2016. Privacy, security, and data protection in smart cities. *Euro. Data Protect. Law Rev.* 2, 1 (2016).
[29] Adel S. Elmaghraby and Michael M. Losavio. 2014. Cyber security challenges in smart cities: Safety, security and privacy. *J. Adv. Res.* 5, 4 (2014), 491–497.
[30] Ali Emadi, Young Joo Lee, and Kaushik Rajashekara. 2008. Power electronics and motor drives in electric, hybrid electric, and plug-in hybrid electric vehicles. *IEEE Trans. Industr. Electron.* 55, 6 (2008), 2237–2245.
[31] Dan Feldman, Cynthia Sung, Andrew Sugaya, and Daniela Rus. 2015. iDiary: From GPS signals to a text-searchable diary. *ACM Trans. Sens. Netw.* 11, 4 (2015), 60.
[32] Institute for Transportation and Development Policy. 2019. Public Bike Data. Retrieved from: http://www.publicbike.net/bikesharing/index/?lang=1.
[33] Nelson Fumo. 2014. A review on the basics of building energy estimation. *Renew. Sustain. Energy Rev.* 31 (2014), 53–60.
[34] Ji Gao, Jack Lanchantin, Mary Lou Soffa, and Yanjun Qi. 2018. Black-box generation of adversarial text sequences to evade deep learning classifiers. Retrieved from: *arXiv preprint arXiv:1801.04354* (2018).
[35] Mario Gerla, Eun-Kyu Lee, Giovanni Pau, and Uichin Lee. 2014. Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds. In *Proceedings of the IEEE World Forum on Internet of Things (WF-IoT'14)*. IEEE, 241–246.
[36] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. 2008. Private queries in location based services: Anonymizers are not necessary. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD'08)*. 121–132.
[37] Noah Goodall. 2014. Ethical decision making during automated vehicle crashes. *Trans. Res. Rec.: J. Transport. Res. Board*2424 (2014), 58–65.
[38] Jayavardhana Gubbi, Slaven Marusic, Aravinda S. Rao, Yee Wei Law, and Marimuthu Palaniswami. 2013. A pilot study of urban noise monitoring architecture using wireless sensor networks. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI'13)*. IEEE, 1047–1052.
[39] Yifeng He, Bala Venkatesh, and Ling Guan. 2012. Optimal scheduling for charging and discharging of electric vehicles. *IEEE Trans. Smart Grid* 3, 3 (2012), 1095–1105.
[40] Abdeltwab M. Hendawi and Mohamed F. Mokbel. 2012. Predictive spatio-temporal queries: A comprehensive survey and future directions. In *Proceedings of the ACM SIGSPATIAL GIS International Workshop on Mobile Geographic Information Systems (MobiGIS*'12).
[41] Martin Hilbert. 2012. Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psych. Bull.* 138, 2 (2012), 211.
[42] Ahsan Raza Khan, Anzar Mahmood, Awais Safdar, Zafar A. Khan, and Naveed Ahmed Khan. 2016. Load forecasting, dynamic pricing and DSM in smart grid: A review. *Renew. Sustain. Energy Rev.* 54 (2016), 1311–1322.
[43] Nicos Komninos, Hans Schaffers, and Marc Pallot. 2011. Developing a policy roadmap for smart cities and the future internet. In *Proceedings of the eChallenges Conference (e-2011)*. IMC International Information Management Corporation.
[44] Raymondus Kosala and Erwin Adi. 2012. Harvesting real-time traffic information from Twitter. *Procedia Eng.* 50 (2012), 1–11.
[45] Jhon Krumm. 2009. A survey of computational location privacy. *Person. Ubiq. Comput.* 13, 6 (2009), 391–399.
[46] Beijing City Lab. 2017. BCL: Beijing City Lab. Retrieved from: https://www.beijingcitylab.com/.
[47] Eun-Kyu Lee, Mario Gerla, Giovanni Pau, Uichin Lee, and Jae-Han Lim. 2016. Internet of vehicles: From intelligent grid to autonomous cars and vehicular fogs. *Int. J. Distrib. Sens. Netw.* 12, 9 (2016).
[48] Wenjia Li, Houbing Song, and Feng Zeng. 2018. Policy-based secure and trustworthy sensing for internet of things in smart cities. *IEEE Internet Things J.* 5, 2 (2018), 716–723.
[49] Thomas Liebig, Nico Piatkowski, Christian Bockermann, and Katharina Morik. 2017. Dynamic route planning with real-time traffic predictions. *Inform. Syst.* 64 (2017), 258–265.

[50] Ruilin Liu, Yu Yang, Daehan Kwak, Desheng Hang, Liviu Iftode, and Badri Nath. 2017. Your search path tells others where to park: Towards fine-grained parking availability crowdsourcing using parking decision models. *Proc. ACM Interact., Mobile, Wear. Ubiq. Technol.* 1, 3 (2017), 78.

[51] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, Fei-Yue Wang et al. 2015. Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intell. Transport. Syst.* 16, 2 (2015), 865–873.

[52] Meiyi Ma, Sarah Masud Preum, and John A. Stankovic. 2017. CityGuard: A watchdog for safety-aware conflict detection in smart cities. In *Proceedings of the 2nd International Conference on Internet-of-Things Design and Implementation.* ACM, 259–270.

[53] Meiyi Ma, Sarah Masud Preum, and John A. Stankovic. 2017. Demo abstract: Simulating conflict detection in heterogeneous services of a smart city. In *Proceedings of the IEEE/ACM 2nd International Conference on Internet-of-Things Design and Implementation (IoTDI'17).* IEEE, 275–276.

[54] Meiyi Ma, S. Masud Preum, W. Tarneberg, Moshin Ahmed, Matthew Ruiters, and John Stankovic. 2016. Detection of runtime conflicts among services in smart cities. In *Proceedings of the IEEE International Conference on (SMARTCOMP'16).* IEEE, 1–10.

[55] Meiyi Ma, John A. Stankovic, and Feng Lu. 2017. Runtime monitoring of safety and performance requirements in smart cities. In *Proceedings of the 1st ACM Workshop on the Internet of Safe Things.* ACM.

[56] Meiyi Ma, John A. Stankovic, and Feng Lu. 2018. CityResolver: A decision support system for conflict resolution in smart cities. In *Proceedings of the International Conference on Cyber Physcial Systems (ICCPS'18).* ACM.

[57] Anzar Mahmood, Nadeem Javaid, and Sohail Razzaq. 2015. A review of wireless communications for smart grid. *Renew. Sustain. Energy Rev.* 41 (2015), 248–260.

[58] Francisco J. Martinez, Chai-Keong Toh, Juan-Carlos Cano, Carlos T. Calafate, and Pietro Manzoni. 2010. Emergency services in future intelligent transportation systems based on vehicular communication networks. *IEEE Intell. Transport. Syst. Mag.* 2, 2 (2010), 6–20.

[59] Antoni Martínez-Ballesté, Pablo A. Pérez-Martínez, and Agusti Solanas. 2013. The pursuit of citizens' privacy: A privacy-aware smart city is possible. *IEEE Commun. Mag.* 51, 6 (2013), 136–141.

[60] Anna Carolina Menezes, Andrew Cripps, Dino Bouchlaghem, and Richard Buswell. 2012. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Appl. Energy* 97 (2012), 355–364.

[61] Fei Miao, Shuo Han, Shan Lin, John Stankovic, Qian Wang, Desheng Zhang, Tian He, and George J. Pappas. 2016. Data-driven robust taxi dispatch approaches. In *Proceedings of the ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS'16).* IEEE, 1–1.

[62] Fei Miao, Shuo Han, Shan Lin, John A. Stankovic, Desheng Zhang, Sirajum Munir, Hua Huang, Tian He, and George J. Pappas. 2016. Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach. *IEEE Trans. Auto. Sci. Eng.* 13, 2 (2016), 463–478.

[63] Atom Mirakyan and Roland De Guio. 2013. Integrated energy planning in cities and territories: A review of methods and tools. *Renew. Sustain. Energy Rev.* 22 (2013), 289–297.

[64] Sirajum Munir, John A. Stankovic, Chieh-Jan Mike Liang, and Shan Lin. 2013. Cyber physical system challenges for human-in-the-loop control. In *Presented as part of the 8th International Workshop on Feedback Computing*. 2013.

[65] Taewoo Nam and Theresa A. Pardo. 2011. Smart city as urban innovation: Focusing on management, policy, and context. In *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance.* ACM, 185–194.

[66] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. 2017. The Mapillary Vistas Dataset for semantic understanding of street scenes. In *Proceedings of the IEEE International Conference on Computer Vision.* 4990–4999.

[67] David Sousa Nunes, Pei Zhang, and Jorge Sá Silva. 2015. A survey on human-in-the-loop applications towards an internet of all. *IEEE Commun. Surv. Tutor.* 17, 2 (2015), 944–965.

[68] City of Copenhagen. 2019. Copenhagen Data. Retrieved from: http://data.kk.dk/dataset.

[69] City of New York. 2017. NYC Open Data. Retrieved from: https://opendata.cityofnewyork.us/.

[70] City of San Francisco. 2019. Data SF. Retrieved from: https://datasf.org/opendata/.

[71] City of Seattle. 2019. Seattle. Retrieved from: https://data.seattle.gov/browse?category=Transportation&provenance=official.

[72] Array of Things. 2019. Array of Things. Retrieved from: http://arrayofthings.github.io/.

[73] City of Zürich. 2019. Stadt Zürich Open Data. Retrieved from: https://data.stadt-zuerich.ch.

[74] NYC OpenData. 2019. OEM Emergency Notifications. Retrieved from: https://data.cityofnewyork.us/Public-Safety/OEM-Emergency-Notifications/8vv7-7wx3.

[75] OpenDataSoft. 2017. Paris Open Data. Retrieved from: http://opendata.paris.fr/.

[76] OpenWeatherMap. 2012–2016. Open Weather Map. Retrieved from: http://openweathermap.org//.

[77] Oliver O'Brien, James Cheshire, and Michael Batty. 2014. Mining bicycle sharing data for generating insights into sustainable transport systems. *J. Transport Geogr.* 34 (2014), 262–273.
[78] Brian Paden, Michal Čáp, Sze Zheng Yong, Dmitry Yershov, and Emilio Frazzoli. 2016. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Trans. Intell. Vehic.* 1, 1 (2016), 33–55.
[79] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. 2014. Sensing as a service model for smart cities supported by internet of things. *Trans. Emerg. Telecomm. Technol.* 25, 1 (2014), 81–93.
[80] Policeone. 2017. Infographic: Twitter Makes Inroads into Police Work. Retrieved from: https://www.policeone.com/social-media-for-cops/articles/6241248-Infographic-Twitter-makes-inroads-into-police-work/.
[81] Sarah Masud Preum, Abu Sayeed Mondol, Meiyi Ma, Hongning Wang, and John Stankovic. 2017. Preclude: Conflict detection in textual health advice. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom'17).* 286–296.
[82] Shaojie Qiao, Changjie Tang, Huidong Jin, Teng Long, Shucheng Dai, Yungchang Ku, and Michael Chau. 2010. PutMode: Prediction of uncertain trajectories in moving objects databases. *Appl. Intell.* 33, 3 (2010), 370–386.
[83] Ho-Kyeong Ra, Asif Salekin, Hee Jung Yoon, Jeremy Kim, Shahriar Nirjon, David Stone, Sujeong Kim, Jong-Myung Lee, Sang Hyuk Son, and John A. Stankovic. 2015. Demo: AsthmaGuide: An ecosystem for asthma monitoring and advice. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems.* ACM, 451–452.
[84] David Rebollo-Monedero, Andrea Bartoli, Juan Hernández-Serrano, Jordi Forné, and Miguel Soriano. 2014. Reconciling privacy and efficient utility management in smart cities. *Trans. Emerg. Telecomm. Technol.* 25, 1 (2014), 94–108.
[85] Miriam Redi, Luca Maria Aiello, Rossano Schifanella, and Daniele Quercia. 2018. The spirit of the city: Using social media to capture neighborhood ambiance. In *Proceedings of the ACM Conference on Human-Computer Interaction* 2 (2018), 144.
[86] Nancy C. Roberts. 2015. *The Age of Direct Citizen Participation.* Routledge.
[87] Haggai Roitman, Jonathan Mamou, Sameep Mehta, Aharon Satt, and L. V. Subramaniam. 2012. Harnessing the crowds for smart city sensing. In *Proceedings of the 1st International Workshop on Multimodal Crowd Sensing.* ACM, 17–18.
[88] Rodrigo Roman, Javier Lopez, and Masahiro Mambo. 2018. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Fut. Gen. Comput. Syst.* 78 (2018), 680–698.
[89] SalamanderLive. 2019. Salamanderlive. Retrieved from: http://www.salamanderlive.com/.
[90] Luis Sanchez, Luis Muñoz, Jose Antonio Galache, Pablo Sotres, Juan R. Santana, Veronica Gutierrez, Rajiv Ramdhany, Alex Gluhak, Srdjan Krco, Evangelos Theodoridis et al. 2014. SmartSantander: IoT experimentation over a smart city testbed. *Comput. Netw.* 61 (2014), 217–238.
[91] Hamidreza Sayarshad, Sepideh Tavassoli, and Fang Zhao. 2012. A multi-periodic optimization formulation for bike planning and bike utilization. *Appl. Math. Mod.* 36, 10 (2012), 4944–4951.
[92] Jesse M. Shapiro. 2006. Smart cities: Quality of life, productivity, and the growth effects of human capital. *Rev. Econ. Stat.* 88, 2 (2006), 324–335.
[93] Pradip Kumar Sharma, Seo Yeon Moon, and Jong Hyuk Park. 2017. Block-VN: A distributed blockchain based vehicular network architecture in smart city. *J. Inform. Proc. Syst.* 13, 1 (2017), 184–195.
[94] smart911. 2016. Smart911. Retrieved from: https://www.smart911.com.
[95] Saleh Soltan, Prateek Mittal, and H. Vincent Poor. 2018. BlackIoT: IoT botnet of high wattage devices can disrupt the power grid. In *Proceedings of the 27th USENIX Security Symposium (USENIX Security'18).* 15–32.
[96] Open Stockholm. 2019. Open Stockholm. Retrieved from: http://open.stockholm.se.
[97] L. Suganthi and Anand A. Samuel. 2012. Energy models for demand forecasting—A review. *Renew. Sustain. Energy Rev.* 16, 2 (2012), 1223–1240.
[98] The Weather Company, LLC. 2017. Wunder Ground Weather. Retrieved from: http://openweathermap.org//.
[99] Zhao-Hua Wang, Hua-Lin Zeng, Yi-Ming Wei, and Yi-Xiang Zhang. 2012. Regional total factor energy efficiency: An empirical analysis of industrial sector in China. *Appl. Energy* 97 (2012), 115–123.
[100] IBM Watson. 2016. IBM Watson IoT Platform. Retrieved from: http://www.ibm.com/internet-of-things/.
[101] Ying Wei, Yu Zheng, and Qiang Yang. 2016. Transfer knowledge between cities. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 1905–1914.
[102] Andrea Zanella, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, and Michele Zorzi. 2014. Internet of things for smart cities. *IEEE Int. Things J.* 1, 1 (2014), 22–32.
[103] Desheng Zhang, Juanjuan Zhao, Fan Zhang, and Tian He. 2015. UrbanCPS: A cyber-physical system based on multi-source big infrastructure data for heterogeneous model integration. In *Proceedings of the ACM/IEEE 6th International Conference on Cyber-Physical Systems.* ACM, 238–247.
[104] Kuan Zhang, Jianbing Ni, Kan Yang, Xiaohui Liang, Ju Ren, and Xuemin Sherman Shen. 2017. Security and privacy in smart city applications: Challenges and solutions. *IEEE Commun. Mag.* 55, 1 (2017), 122–129.

[105] Yu Zheng. 2015. Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 29.

[106] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-Air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1436–1444.

[107] Yu Zheng, Xiuwen Yi, Ming Li, Ruiyuan Li, Zhangqing Shan, Eric Chang, and Tianrui Li. 2015. Forecasting fine-grained air quality based on big data. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2267–2276.

[108] Minqi Zhou, Rong Zhang, Wei Xie, Weining Qian, and Aoying Zhou. 2010. Security and privacy in cloud computing: A survey. In *Proceedings of the 6th International Conference on Semantics Knowledge and Grid (SKG'10)*. IEEE, 105–112.

[109] Bing Zhu, Qixing Huang, Leonidas Guibas, and Lin Zhang. 2013. Urban population migration pattern mining based on taxi trajectories. In *Proceedings of the 3rd International Workshop on Mobile Sensing: The Future, Brought to You by Big Sensor Data*.

Received February 2018; revised June 2019; accepted June 2019