Architecting Smart City Digital Twins: Combined Semantic Model and Machine Learning Approach

Mark Austin[1]; Parastoo Delgoshaei[2]; Maria Coelho[3]; and Mohammad Heidarinejad, M.ASCE[4]

Abstract: This work was motivated by the premise that next-generation smart city systems will be enabled by widespread adoption of sensing and communication technologies deeply embedded within the physical urban domain. These technological advances (e.g., sensing, processing, and data transmission) are what makes smart city digital twins possible. This paper explores approaches and challenges in architecting and the operation of smart city digital twins. A smart city digital twin architecture is proposed that supports semantic knowledge representation and reasoning, working side by side with machine learning formalisms, to provide complementary and supportive roles in the collection and processing of data, identification of events, and automated decision-making. The semantic and machine learning sides of the proposed architecture are exercised on a problem involving simplified analysis of energy usage in buildings located in the Chicago Metropolitan Area. **DOI: 10.1061/(ASCE)ME.1943-5479.0000774.**

Author keywords: Smart city; Digital twin; Semantic modeling; Machine learning.

Introduction

Problem Statement

Rapid urbanization places additional demands on cities that are constrained to operate with limited resources. Long-term solutions to this bottleneck are driving sociotechnological transformation of cities into smart cities, which includes the adoption of new technologies imitating human intelligence. One of the important outcomes of advances in computing and communications—the fifth generation of cellular mobile communications (5G) and Internet of Things (IoT)—over the past few decades is the way in which they have opened doors to the replacement of aging urban infrastructure with new types of urban systems comprising physical networks connected to cyber components (data, information, software) for decision-making (Preuss 2017). Looking forward, these next-generation urban systems will be defined by superior levels of performance, new forms of functionality, transparency in allocation of resources, and good economics over long time horizons. While city dwellers enjoy the benefits that these technological advances afford, systems engineers and urban planners are faced with a multitude of new design challenges. These challenges arise from the presence of heterogeneous content (multiple disciplines, multiple types of data and information, multiple systems of sensing and measurement), network structures that are spatial, multilayer, interwoven, and dynamic, behaviors and control that are distributed and concurrent, and interdependencies among coupled urban subsystems comprising physical, cyber, organizational, and social domains (Rinaldi et al. 2001).

Solutions to these challenges are complicated by the need to address the myriad of questions faced by urban stakeholders, while at the same time taking into account the unique physical, economic, social, and cultural characteristics of individual cities. From an operations standpoint, two basic questions are: (1) What strategies of day-to-day operation lead to high levels of performance and (2) How do different cities respond to and recover from human-caused and natural disasters? The authors believe that high levels of situational awareness (i.e., understanding how a city is actually used) are a prerequisite to improving day-to-day operations. Similarly, from a long-term planning perspective, accurate estimation of future demands on limited resources is essential for achieving healthy and sustainable urban behavior (Ramaswami et al. 2018). In an improvement on decision-making procedures from a bygone era, answers to these questions can now be based on data-driven approaches to measuring urban performance and to identifying trends and spatiotemporal patterns in city behavior.

The distributed and multidomain nature of urban systems means that no single decision maker knows all of the information known to all of the other urban decision makers. In order for the communication (and dependency) relationships among participating urban domains to occur in an orderly and predictable way, designers need to pay attention to the boundaries and interactions connecting urban domains (Selberg and Austin 2008). This observation suggests that a system for smart management of urban behaviors will follow an architecture along the lines illustrated in Fig. 1, with analytical tools providing strategies for real-time control of multidomain behaviors and planning of actions in response to events distributed throughout the urban environment. Formal approaches to spatiotemporal reasoning are needed to ensure that actions (the results of decisions)

[1]Associate Professor, Dept. of Civil and Environmental Engineering, and Institute for Systems Research, Univ. of Maryland, College Park, MD 20742 (corresponding author). ORCID: https://orcid.org/0000-0001-8259-4794. Email: austin@isr.umd.edu

[2]Research Engineer, Mechanical Systems and Control Group, Engineering Laboratory, National Institute of Technology, 100 Bureau Dr., Stop 8631, Gaithersburg, MD 20899. Email: parastoo.delgoshaei@nist.gov

[3]Graduate Student, Ph.D. Program in Civil Systems, Dept. of Civil and Environmental Engineering, Univ. of Maryland, College Park, MD 20742. Email: mecoelho@terpmail.umd.edu

[4]Assistant Professor, Dept. of Civil, Architectural, and Environmental Engineering, Alumni Memorial Hall, Room 204, Illinois Institute of Technology, 3201 South Dearborn St., Chicago, IL 60616. ORCID: https://orcid.org/0000-0002-3064-825X. Email: muh182@iit.edu

Note. This manuscript was submitted on March 14, 2019; approved on December 12, 2019; published online on April 8, 2020. Discussion period open until September 8, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Management in Engineering***, © ASCE, ISSN 0742-597X.

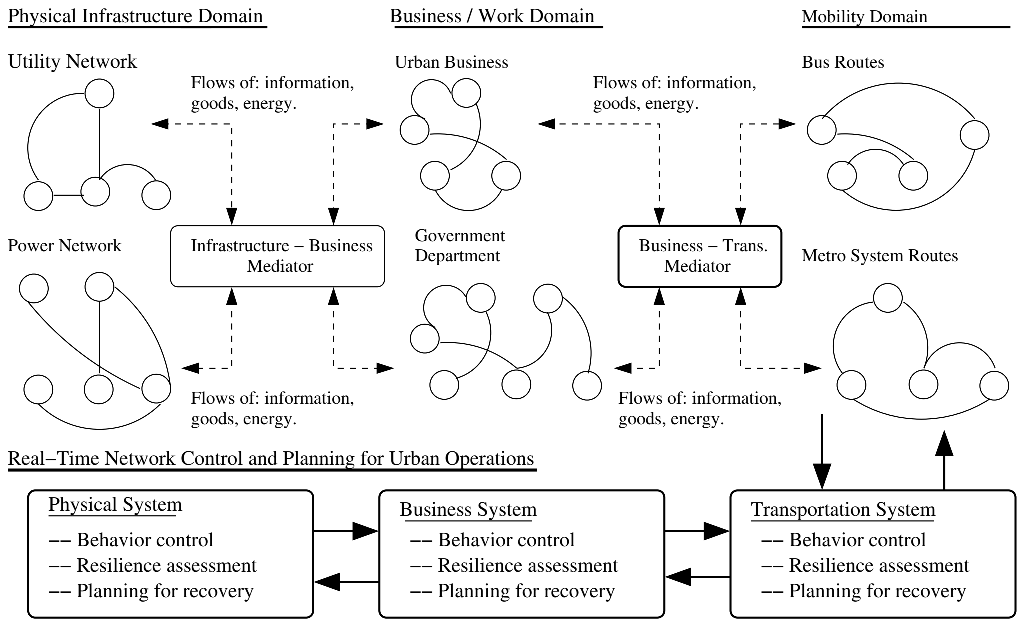

Fig. 1. Architecture for multidomain urban behavior modeling and computational support for real-time control of interactions—flows of data, information, people, goods, and energy—among urban domains. (Adapted from Coelho et al. 2017.)

occur in the right place and at the right time. These approaches need to operate across multiple spatial and temporal scales. It has also been observed that, unlike standard cradle-to-grave product life cycle models, cities evolve in response to economic opportunity and external threats. Formal support is needed for the short- and long-term managed evolution of these opportunities and challenges.

Scope and Objectives

This work was motivated by the premise that next-generation urban systems will be enabled by widespread adoption of sensing and communication technologies deeply embedded within the physical urban domain. These technological advances will allow for the development of smart city digital twins (Lee et al. 2015; Savisalo et al. 2018). A smart city digital twin is defined here as a cyber component that mirrors the physical urban system through real-time monitoring and synchronization of urban activities. Appropriate software and algorithms will work to provide superior levels of urban performance (e.g., in urban mobility, energy efficiency), urban planning (e.g., zoning), and resilience (e.g., through strategies of control and risk management).

This paper explores approaches and challenges in architecting and the operation of smart city digital twins. Realization of this opportunity is complicated by the reality that within the world of model-centric engineering, present-day use of artificial intelligence (AI) and machine learning (ML) technologies is fragmented and at a crossroads. During the past decade engineering researchers in AI have tended to focus on the comprehensive development of ontologies for a domain or development activity and their extension from common core ontologies (e.g., geospatial, time, actors, events) and higher-level basic formal ontologies (Arp et al. 2015). Far less attention has been given to the development of rules associated with ontologies and consideration of the ways in which ontologies and rules can work together to respond to events and support decision-making. At the same time, ML algorithms provide comprehensive support for the classification, clustering, and identification of association relationships and anomalies in streams of real-world data. Remarkable advances (Cai et al. 2017) in ML algorithms include the ability of a machine to learn the structure of large-scale graphs and their attributes. And yet, ML techniques struggle to explain the rationale for decision-making, a task that multidomain semantic modeling and rule-based reasoning can complete with ease.

With these challenges in mind, it is proposed here that the best pathway forward for smart city digital twin design is with architectures that support AI and ML formalisms working side by side as a team, providing complementary and supportive roles in the collection and processing of data, identification of events, and automated decision-making. As illustrated in Fig. 2, the authors envision a so-called city operating system that provides city stakeholders (residents, businesses, planners, and engineers) with enhanced levels of situational awareness and decision-making support for the management of urban infrastructure and services embedded in the space-time domain. The knowledge representation and reasoning (KRR) component entails domain and meta-domain data, ontologies, and rules that can dynamically respond to events. KRR encapsulates the data and represents it with reduced dimensionality, and ML will be used to classify data into collections and learn about cause-and-effect relationships embedded in the data. The proposed approach builds upon the authors' recent work in semantic modeling for urban system of systems (Austin et al. 2015; Coelho et al. 2017) and exploration of a combined semantic and ML approach to the monitoring of energy consumption in buildings (Delgoshaei et al. 2018).

The remainder of this paper proceeds as follows. First, related work in the areas of semantic modeling and rule-based reasoning, ML and data mining techniques, and digital twins is covered. Next, the proposed smart city digital twin architecture is introduced. This section describes an architectural template where semantic modeling and rule-based reasoning work alongside technologies for ML. Extensions of this template are then proposed to highly multidisciplinary urban domains, and application of the proposed methodology to simplified analysis of energy usage is demonstrated in buildings located throughout the Chicago Metropolitan area. The paper closes with a discussion and conclusions.

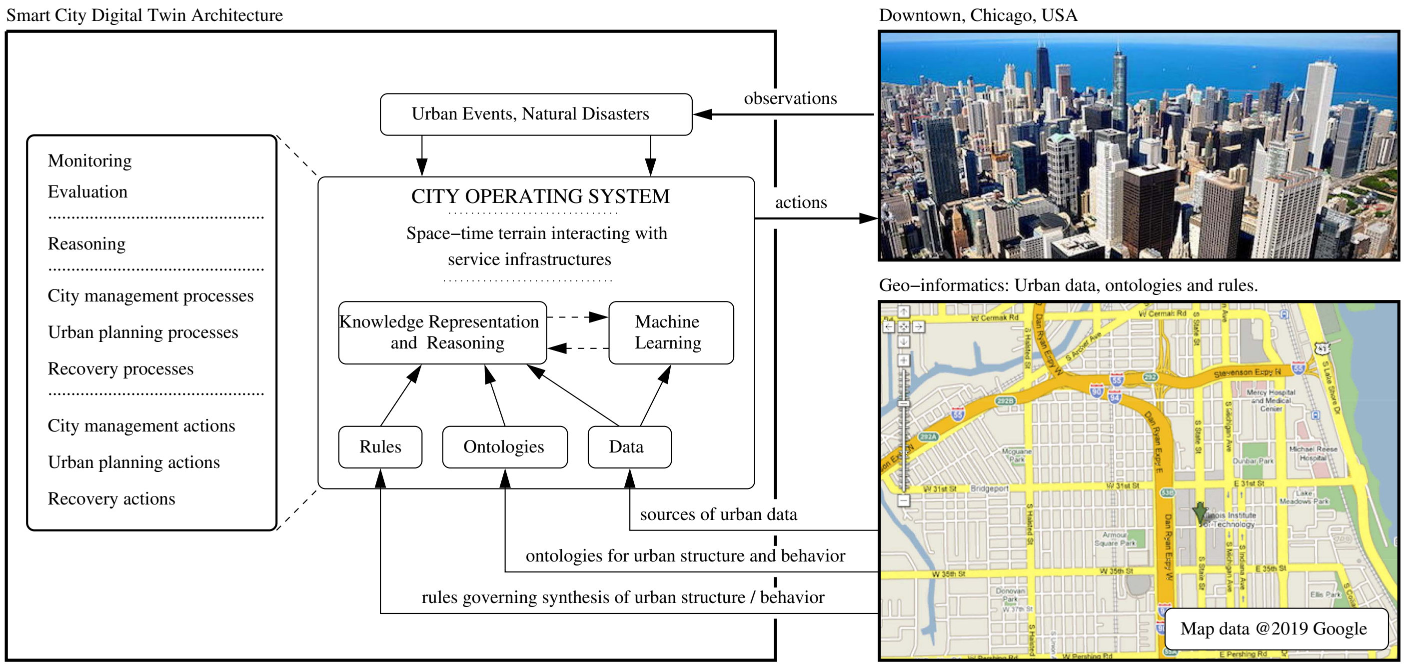

Fig. 2. Smart city digital twin architecture and operating system view of smart city behaviors, management of city and urban planning processes and actions, and restoration of operations in response to disruptions [Downtown Chicago Image: Andrew Horne, under Creative Commons-BY_2.0 license (https://creativecommons.org/licenses/by/2.0/); Map data (c) 2019 Google].

Related Work

This section discusses related work in semantic modeling with ontologies and rule-based reasoning, supervised and unsupervised ML, and digital twins.

Semantic Modeling and Rule-Based Reasoning

A long-standing tenet of the authors' work (Casey and Austin 2002; Austin et al. 2006) has been that methodologies for strategic approaches to urban design will employ semantic descriptions of application domains. These approaches use ontologies and rule-based reasoning to enable validation of requirements and communication (or mappings) among multiple disciplines. The top section of Fig. 3 complements Figs. 1 and 2 and pulls together the different pieces of the proposed architecture for distributed system behavior modeling with ontologies, rules, mediators, and message-passing mechanisms. On the left-hand side, the textual requirements are defined in terms of mathematical and logical rule expressions for design rule checking. Urban domain models correspond to a multitude of graph structures and composite hierarchy structures for the system structure and system behavior. Behaviors are associated with components. Discrete behavior will be modeled with finite state machines. Continuous behaviors will be represented as the solution to ordinary and partial differential equations. Ontology models and rules glue the requirements to the engineering models and provide a platform for simulating the development of system structures, adjustments to system structure over time, and system behavior. The urban building-block level of Fig. 3 shows illustrative examples of ontologies and rules for the definition of building block elements (e.g., requirements, networks, sensors) used in larger multidomain semantic models. The meta-domain section of Fig. 3 shows ontologies and rules (e.g., space, time, physical units) that can be imported into and are fundamental to the representation of concepts and reasoning across all domains.

The development of Semantic Web technologies (Berners-Lee et al. 2001) over the past two decades has resulted in the ability for machines to access and share information, thereby supporting automated discovery of new knowledge (Hendler 2001). For the semantic modeling of multidomain urban settings, the use of Semantic Web technologies for rule checking has several key benefits:

- Rules that represent policies are easily communicated and understood.
- Rules retain a higher level of independence than logic embedded in systems.
- Rules separate knowledge from its implementation logic.
- Rules can be changed without changing the source code or the underlying model.

A rule-based approach to problem solving is particularly beneficial when the application logic is dynamic (i.e., where a change in a policy needs to be immediately reflected throughout the application) and rules are imposed on the system by external entities. Rules can be developed to resolve situations of conflict or competing objectives—such strategies use notions of fairness to prevent deadlocks in the system operation. All three of these conditions apply to the design and management of urban systems. Together, the eXtensible Markup Language (XML), resource description framework (RDF), and Web Ontology Language (OWL) allow for the implementation of reasoning that can prove whether or not assertions are true or false. These tools need to operate in (almost) real time, and as such, description logic requires extensions to make them computationally decidable (Baader et al. 2008). The reader is referred to Petnga and Austin (2016) for a detailed summary of description logic concepts. Finally, this work employs Apache Jena (2016), an open-source Java framework for building Semantic Web and linked data applications. Jena provides application programming interfaces (APIs) for developing code that handles RDF, OWL, and SPARQL (support for query of RDF graphs). The result is a formal rule-based approach to the processing of incoming events, reasoning, and adjustment to the structure and properties of semantic graphs.

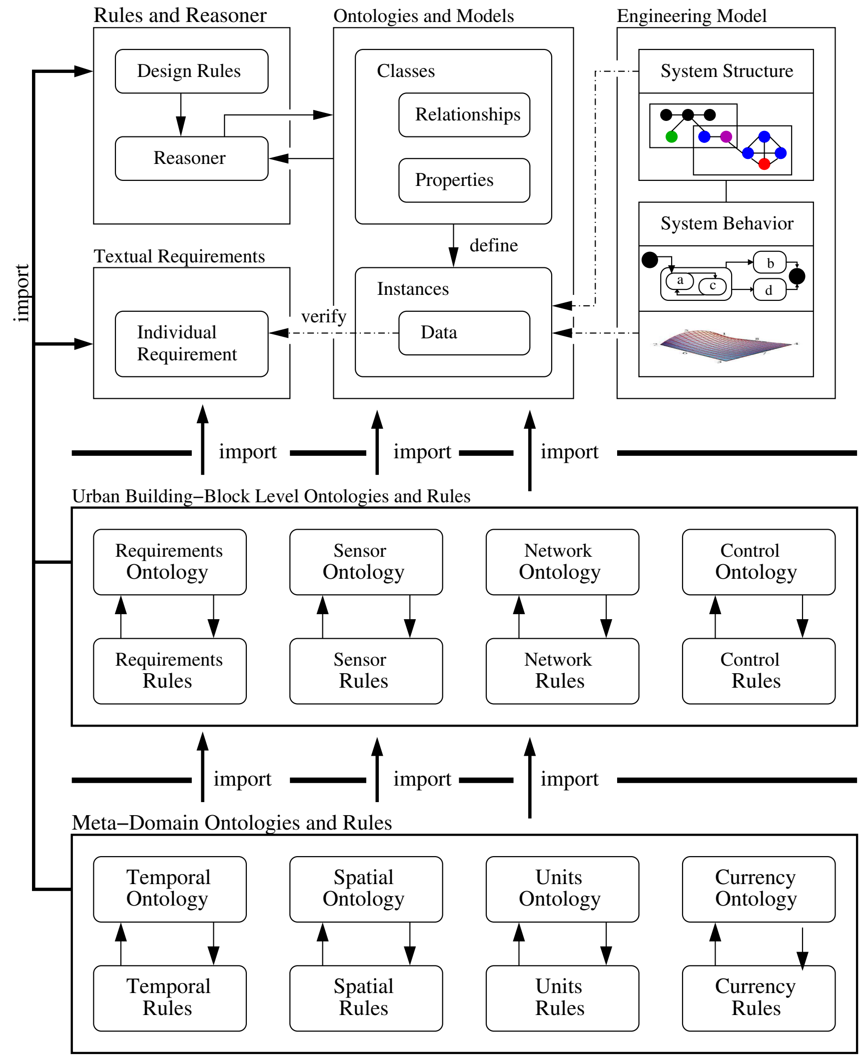

Fig. 3. Framework for implementation of semantic models using ontologies, rules, and reasoning mechanisms.

A detailed discussion on the use of ontologies in urban development projects can be found in Falquet et al. (2005). The scope of contributions includes the development of ontologies for the geographic information sector, modeling interconnections (mediators) among urban models, and describing urban mobility processes. This work also includes the development of ontologies for the Geography Markup Language (GML) and CityML, the XML markup language for cities.

Machine Learning and Data Mining Techniques

Machine learning and data mining techniques learn about a system's structure and behavior, thereby providing insight—sets of patterns or expectations—into the underlying (raw) data, and support decision-making and prediction of future system states. These methods are used to solve complex engineering applications that entail a large number of independent parameters and nonlinear interdependencies that cannot be easily modeled from first principles. For the purposes of this paper, understanding the built environment in large-scale urban areas and spatiotemporal patterns of building energy consumption are among these applications.

ML algorithms can use supervised or unsupervised learning. Supervised learning typically encompasses two steps: training and prediction. Data sets are divided into training data sets and testing data sets. The training step allows identification of the decision model that provides the dependency of the target (predicted variable) on the features (impacting variables). In the next step, the decision model is applied to the testing data sets, and then the effectiveness of the prediction performance of the model can be calculated. Supervised learning, such as the nearest-neighbor algorithm, requires labeled data sets (e.g., the data are labeled with the correct answer), a process that can be computationally expensive. This algorithm makes predictions on new data points based on their proximity to the points in the training set. Supervised ML techniques also include the discovery of binary decision trees.

The goals of unsupervised learning algorithms (e.g., the k-means clustering algorithm) are to model and identify the underlying structure or patterns in a data set when no correct answers (labels) are provided. Semisupervised learning methods fall between the strategies of supervised and unsupervised learning and employ combinations of labeled and unlabeled data. First, the unsupervised method is used to identify patterns, and then supervised learning is used to make the best predictions for the unlabeled data using the labels generated by unsupervised learning. The prediction decision model is then tested on labeled data. This semisupervised technique can address a wide range of engineering applications, including building energy performance and procedures needed for the recovery of operations after disruptions in an urban environment.

Digital Twins

A digital twin is a cyber (or digital) representation of a system that mirrors its implementation in the physical world through real-time monitoring and synchronization of data associated with events. The associated software and algorithms work to provide superior levels of performance, strategies for avoiding unnecessary down time, and support for long-term planning. Digital twins may also avoid the need to tinker with the physical system itself. The digital twin concept dates back to the 2000–2010 era (Glaessgen and Stargel 2012) and was initially proposed as a way to support the design and operation of air vehicles for NASA. Since then the range of potential applications has expanded to include automotive components, manufacturing processes, power plants, design of networks of wind turbines, and smart cities (Lee et al. 2015; Mohammadi and Taylor 2018; Negri et al. 2017; Privat et al. 2019; Kaewunruen et al. 2018; Tahmasebinia et al. 2019). Common key elements for successful digital twins are cloud computing, IoT, and semantic modeling (Alam and Saddik 2017; M2M 2015; Nie et al. 2019; Tao et al. 2018). Within the systems engineering community, Siemens now sees digital twins as the successor to procedures for model-based systems engineering (Boschert et al. 2018).

Note that the concept of building a digital replica of a physical object or space is also found in building information modeling (BIM), parametric three-dimensional (3D) computer-aided design (CAD) technologies, and processes in the architecture, engineering, and construction (A/E/C) industry. Like digital twins, BIM became popular in the early 2000s by providing contractors on large projects a central point of building reference in a 3D digital, visual, and quantitative model of buildings. When coupled with the integration of work practices among architects, engineers, fabricators, and contractors, BIM practice paradigms evolve cumulatively along a trajectory from visualization to coordination to analysis and, finally, to supply chain integration (Taylor and Bernstein 2009). Adopting BIM practice can lead to tremendous improvements in project productivity, such as lowering the risk of projects through a reduction in errors, better time lines, and budget management (Jin et al. 2017). However, this physical information model is suited for buildings during concept and design phases, not ones occupied and utilized daily. Another way of stating this distinction is as follows: BIM is just a picture of what the real-world object should be, whereas a digital twin is a digital replica of an existing and operational asset. As such, the digital twin for a building can provide insight into the current state of building subsystems and how they are impacted by occupant behavior and other events.

Proposed Digital Twin Architecture

An idealized implementation of a digital twin consists of a model of physical objects, data stored with the objects, and a unique one-to-one correspondence between individual objects and an ability to monitor and synchronize their state and behavior (Shetty 2017). For smart city digital twins (Lee et al. 2015; Savisalo et al. 2018) the path to an implementation is complicated by a multitude of concerns spanning the physical, cyber, social, and natural domains (Figs. 1 and 3), as well as difficulties in defining semantics and rules for their interaction. Further challenges arise when knowledge of the urban structure and behavior is incomplete and considerable uncertainty exists in the spatial and temporal nature of future urban events. Looking forward, smart city digital twins will need to deal with these challenges. The authors propose that one way of doing this is to have them iterate on a process of continual learning that includes progressive refinement of semantic models and data-driven discovery of cause-and-effect relationships and hidden patterns among elements of the participating domains. The building block for this work is an architectural template that integrates semantic modeling with ML techniques and defines their roles and interactions. It is demonstrated how the template might be expanded to include the range of data sources and measures for assessment found in a real-world city.

Semantic Modeling and Machine Learning Architectural Template

Fig. 4 shows the proposed architectural template for a combined multidomain semantic modeling and ML approach to the implementation of smart city digital twins. Instead of creating a small number of all-encompassing ontologies and associated rules, the goal here is to put the development of data, ontologies, and rules on an equal footing and create architectural templates for a specific domain or design concern (a convenient name is the data-ontology-rule footing). This approach to semantic model development forces developers to think about the chain of dependency relationships among the data, ontologies, and rules and to provide the data needed to support decision-making—rules require data and object properties from the ontologies, which in turn require data from the data models shown along the right-hand side of Fig. 4. Semantic graph models will be populated with individuals (i.e., instances of real-world data) by visiting (a software design pattern) the relevant data models and gathering the data and object properties relevant to the application at hand. Rules can be developed for the verification of semantic properties (e.g., to verify that a specific data property has been initialized) and for reasoning with data sources and incoming events, possibly from a multiplicity of domains. Implementation of the latter leads to semantic graphs that can dynamically adapt to incoming events (e.g., weather).

The proposed architectural template also employs ML techniques and three styles of learning to gain insight into the data. Semantic feature engineering (Zheng and Casari 2018) begins with raw data and then uses domain knowledge to create features that are used in ML algorithms, which supports the organization (and simplification) of rules. Tree-based classification algorithms involve the identification of binary rules (true/false) and branching based on numerical and temporal data that maximize the likelihood of prediction of a target variable. Clustering algorithms identify groups of things that belong together. Knowledge of these groups is useful for architecting ontologies and related data and object properties in a manner that is consistent with the underlying data. Association algorithms look for rules that strongly associate different data attribute values, possibly spanning domains, and then use this information to develop cross-domain rules for desirable behavior. For example, in the smart city application, an association rule might uncover the observation that the majority of office spaces are in a specific area of a city.

Combined Semantic Modeling and Machine Learning of Urban Domains

Fig. 5 builds upon Figs. 3 and 4 and shows an abbreviated digital twin architecture for a combined semantic modeling and ML approach to the monitoring and real-time management of smart city operations. One key benefit in organizing data, ontologies, and rules into template structures is that it opens doors to the systematic assembly of system-level models of urban behavior and properties through the composition of smaller template models. A fully

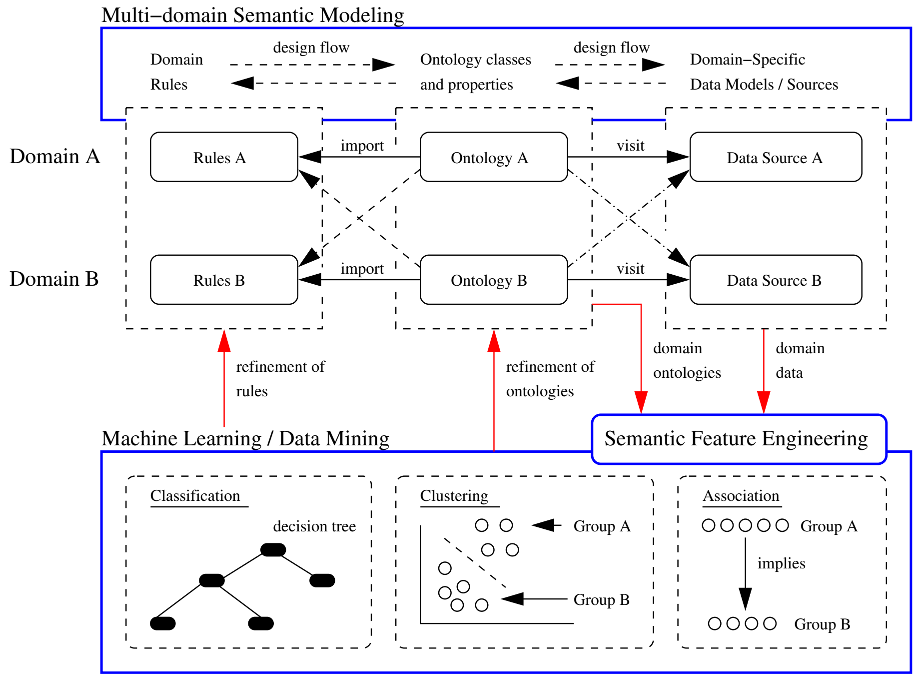

Fig. 4. Simplified architectural template for combined semantic and ML modeling used in digital twins.

implemented smart city digital twin (see the rows of blocks in Fig. 5) covers the physical and cyber infrastructure, business operations and supply chains, urban services and land use, and events generated by external events such as weather. Most of these models are embedded in spatiotemporal terrain. Domain-specific ontologies and rules are populated with data from a wide range of sources: examples include Open Street Map (OSM), GML and CityGML for the population of urban network ontologies (Falquet et al. 2005; OSM 2019), online weather servers for the population of weather ontologies, city hall data for the development of urban policies and regulations, and census data for population demographics. With the advent of smart devices and network connectivity, urban operations can now be monitored with data that are crowd-sourced. An early example of this is the acoustic sensor network currently in development in New York (Bello et al. 2019).

The lowermost section of Fig. 5 shows how ontologies are imported into a semantic graph, rules are loaded into a reasoner, and semantic graphs are attached to reasoners to process incoming events and potentially trigger transformations in the semantic graph. It is important to note that support for cross-domain reasoning and the inference of new knowledge occurs through the development and execution of ontology namespaces from multiple urban domains, as well as the multidomain ontologies for space and time shown along the bottom of Fig. 3. The authors believe that this framework is sufficient for the development of reasoning procedures that can (1) ensure actions taken on an object or urban subsystem in response to an event occur at the right place and at the right time, (2) provide insight into the propagation of domain interactions and, in the longer term, the emergence of patterns in city behavior, and (3) support a validation mechanism that insures the semantic model is a true representation of the real world.

Case Study Problem

This section presents a case study problem where urban data from the publicly available Chicago Energy Benchmarking data set (Chicago 2019) is analyzed from data mining (ML) and semantic modeling perspectives. This data set is part of Chicago's Open Data Portal, an ongoing effort to make government more responsive, transparent, and cost-effective (Goldsmith and Crawford 2014). In 2016, the building sector in the US accounted for approximately 40% of total energy consumption (EIA 2016), suggesting that with advances in technology over the past few decades, opportunities exist to tackle energy use and design near zero-energy buildings (Shrivastava and Chini 2016; Hong et al. 2012; Shrestha and Kulkarni 2013). The purpose of the case study is twofold. The first goal is to understand the extent to which the data set supports (1) the classification of buildings at the neighborhood (or community area or zip code) level, and (2) the prediction of building energy use as a function of data set attributes, allowing for the design of building ontologies and rules on the semantic side of the problem. The second purpose is to understand the extent to which the data set supports (or, conversely, does not support) hypotheses that an engineer might have. An engineer might wonder, for example, whether, on average, buildings close to Lake Michigan consume quantities of energy significantly higher than the average values. One might also wonder whether buildings north of the city center consume more energy than those to the south. Knowledge of such trends, even if they are subject to uncertainty, are useful for stakeholders concerned with energy consumption (e.g., building managers, energy companies, city managers). Preliminary insight into these issues is possible with spreadsheet-like computations. Beyond that, the Weka (Witten et al. 2017) tool is used in data mining experiments and Apache Jena for the construction of semantic graph models and rules.

Chicago Energy Benchmarking Data Set

The Chicago Energy Benchmarking data set contains 2 years of self-reported building annual energy consumption data. The buildings in this database are commercial (e.g., hotels, retail stores, office buildings), residential (e.g., multifamily housing units), and

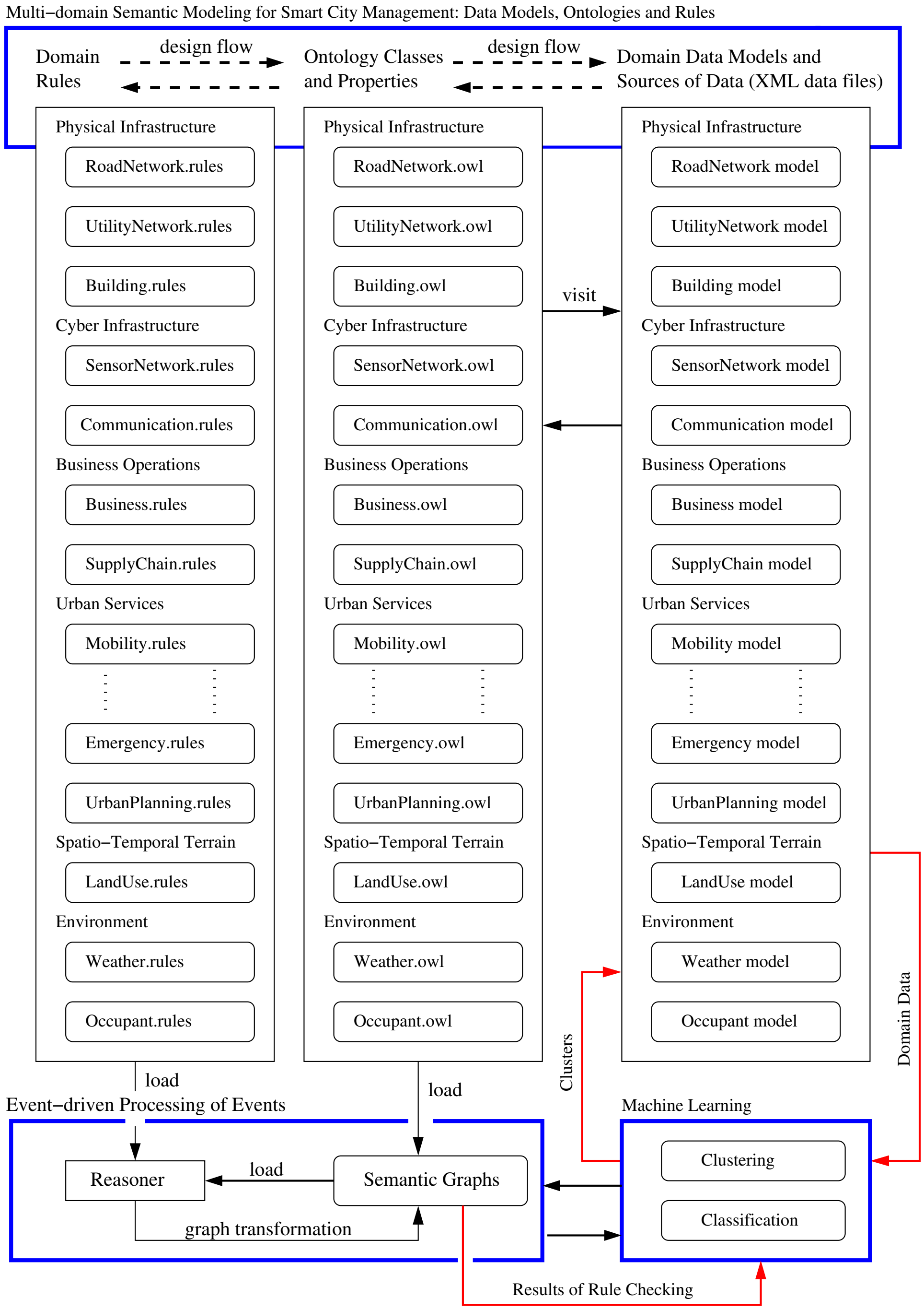

Fig. 5. Abbreviated digital twin architecture for combined semantic and ML approach to monitoring and real-time management of smart city operations. Data, ontologies, and rules are placed on equal footing and developed for multiplicity of domains.

civic and municipal buildings (e.g., performing arts, hospitals, libraries) with a gross floor area exceeding 4,645 m^2 (50,000 ft^2).

Fig. 6 shows the spatial distribution of buildings and zip codes in the Chicago Energy Benchmarking data set. The left-hand side shows a bird's-eye view of zip codes and the distribution of buildings in the area of study. The right-hand side of Fig. 6 focuses on the downtown area and, in particular, zip code 60616, which will be used to illustrate the results of learning in the computational experiments. Each building is defined by 26 attributes; for the present purposes, the most important are (1) property name and address, (2) zip code, (3) community area, (4) primary property type, (5) gross floor area ($ft^2 = 0.0929\,m^2$), (6) year built, (7) age, (8) electricity usage (kBtu = 0.293 kWh), (9) natural gas usage (kBtu = 0.293 kWh), (10) site energy utilization index (EUI) ($kBtu/ft^2 = 3.14\,kWh/m^2$), (11) source EUI ($kBtu/ft^2 = 3.14\ kWh/m^2$), and (12) location (i.e., latitude and longitude). Two factors that complicate the use of data set attributes are the codependency between the attributes and lack of precision. An example of the former is Attributes 6 and 7, which are coupled by a simple relation.

Preliminary Analysis of Building Energy Consumption

State-of-the-art practice for the assessment of EUI in buildings is defined by EnergyStar (2018), which provides median scores for site and source EUI for all types of buildings. The national median

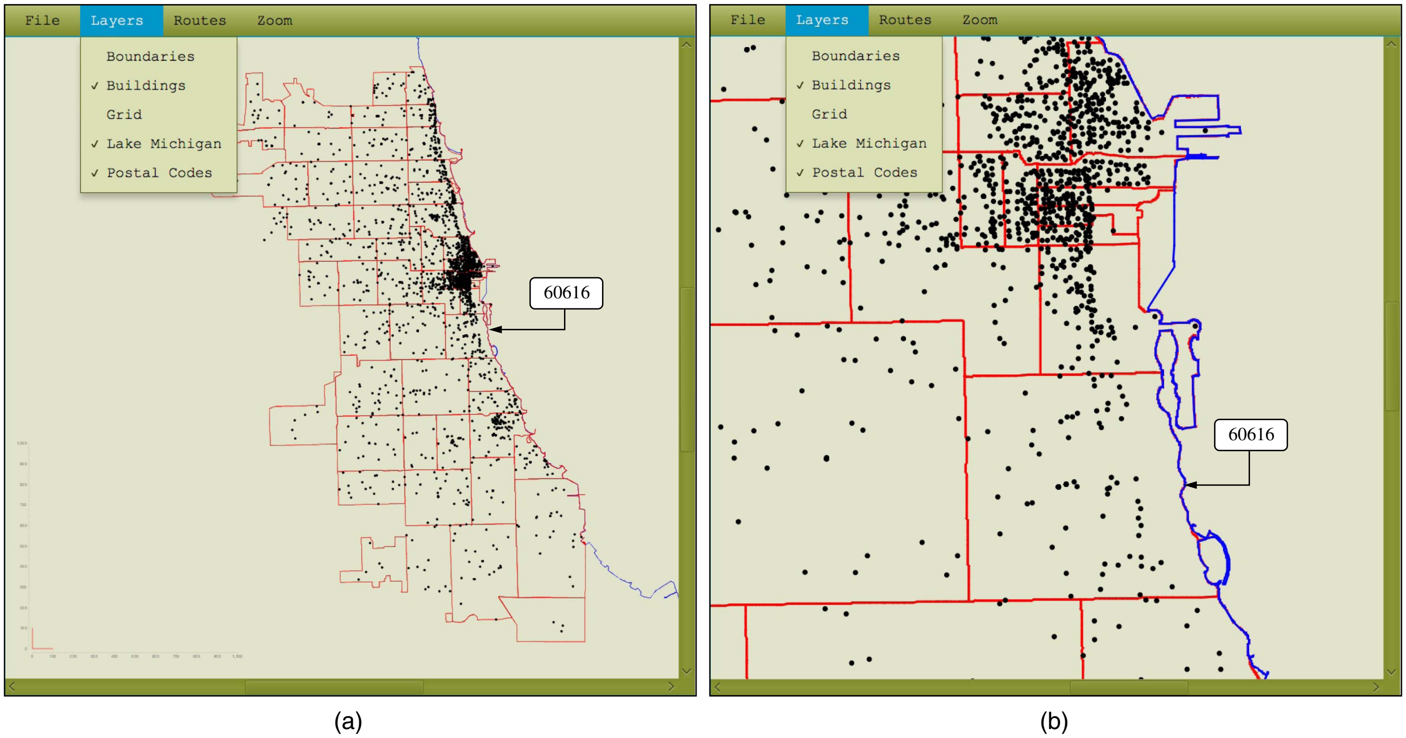

Fig. 6. Spatial distribution of buildings and zip codes represented in Chicago Energy Benchmarking data set. Left-hand side: bird's-eye view of 55 zip code regions and buildings. Right-hand side: focus on downtown area and region for zip code 60616.

data indicate, for example, that hospitals have larger site EUI values compared to schools. While this framework provides building portfolio managers with a means to assess how well individual buildings are performing relative to their peers, median values are a single data point and ignore the scatter in data across a spectrum of buildings of the same type. As a first step toward addressing this shortcoming, Table 1 is a summary of electricity and gas usage (kBtu) in various kinds of buildings located in the Chicago Metropolitan Area. On a per-entity basis, the four most significant categories of electricity and gas consumption stem from colleges and universities (4.3% and 5.76%), hospitals (7.59% and 9.16%), multifamily housing units (26.46% and 48.59%), and office buildings (31.09% and 9.25%). Accordingly, from this point on, building types outside these four areas were removed from the scope of this case study.

Table 2 shows the distribution of source EUI organized across five zones for hospitals, multifamily housing, office, and university buildings (1,788 buildings in total). The buildings are sorted by source EUI value and then assigned to one of five zones. EUI Zone 1 (0%–20%) covers the interval $[12.1, 103.6](\mathrm{kBtu/ft^2})$, $[38{,}325.3](\mathrm{kWh/m^2})$; EUI Zone 2 (20%–40%) covers the interval $[103.7, 130.3](\mathrm{kBtu/ft^2})[325.6{,}409.1](\mathrm{kWh/m^2})$; EUI Zone 3 (40%–60%) covers the interval $[130.4{,}152.7](\mathrm{kBtu/ft^2})[409.4{,}479.4](\mathrm{kWh/m^2})$; EUI Zone 4 (60%–80%) covers the interval $[152.8, 189.3](\mathrm{kBtu/ft^2})[479.8, 576.9](\mathrm{kWh/m^2})$; and EUI Zone 5 (80%–100%) covers the interval $[190.2, 970](\mathrm{kBtu/ft^2})[597.2, 3045.8](\mathrm{kWh/m^2})$. Hospitals, university buildings, and offices constitute 25% of the building inventory and are relatively heavy—at least 50% of the buildings are in Zone 5—consumers of energy. Otherwise, the building inventory is dominated by (1,331 out of 1,788) multifamily housing, which has source EUI values that are (roughly speaking) uniformly distributed across the five zones. Table 3 shows the same statistics for buildings located within zip code 60616. Even within a single zip code, the energy usage of multifamily housing units covers a wide range of values, indicating that mean values of energy usage by themselves provide an incomplete picture in the assessment of building energy usage.

Semantic Feature Engineering and Preliminary Semantic Models

Semantic feature engineering (Fig. 4) is defined here as the process of using a combination of domain knowledge and inference rules to infer features that can serve as meaningful input to data mining algorithms (see arrows pointing downwards from multi-domain semantic modeling to semantic feature engineering) for tree-based classification, clustering, and extraction of association relationships. Similarly, the arrows pointing upwards from semantic feature engineering to multi-domain semantic modeling correspond to ML algorithms providing feedback for the specification and organization of digital twin ontologies and, potentially, development and simplification of urban rules. With this framework in place, semantic rules can extract features from a data set and synthesize new implicit information. For example, by comparing the latitude and longitude for a building with geographic profiles of the lake boundary, a simple semantic rule can create a boolean data property indicating whether or not (true or false) the building is close to the lake.

For the purposes of this case study, the data parameters will be mined from a combination of (1) attributes found in the City of Chicago Building Energy Consumption Benchmarking data set and (2) classes and data properties found in domain ontologies for buildings, land use, and urban planning (Fig. 7). The role of these domains in the larger urban picture is shown in Fig. 5. The purpose of the building ontology and associated semantic rules is to provide semantic information and perspective regarding the aspects of building energy performance and urban planning, in this case a small number of details about zoning constraints and definitions. Such constraints can ensure, for example, that an industrial space will not be built next

Table 1. Summary of analysis results for building energy consumption across Chicago Metropolitan Area in 2016

Property type	No electricity	Entity use (kBtu = 0.293 kWh)	Percentage	Natural gas use (kBtu = 0.293 kWh)	Percentage
Adult education	3	28,692,102.0	0.10	18,564,922.6	0.07
Auto dealership	4	27,273,443.2	0.09	21,831,307.5	0.08
Bank branch	3	13,061,369.1	0.05	12,537,212.1	0.04
College/University	87	1,238,968,865.6	4.31	1,642,444,340.9	5.76
Convention center	1	485,705,024.0	1.69	325,947,797.7	1.14
Courthouse	2	66,813,357.5	0.23	63,930,413.0	0.22
Education	3	14,283,524.6	0.05	21,849,218.7	0.08
Enclosed mall	5	212,613,034.3	0.74	136,790,519.6	0.48
Financial office	4	340,885,369.2	1.18	198,475,181.2	0.70
Fitness center	9	81,582,735.1	0.28	84,346,836.6	0.30
Hospital	30	2,184,696,752.3	7.59	2,610,466,870.4	9.16
Hotel	75	1,534,256,336.5	5.33	1,690,997,910.8	5.93
Indoor arena	1	113,634,101.0	0.39	8,970,953.0	0.03
K-12 school	387	1,361,021,876.3	4.73	2,078,021,816.5	7.29
Laboratory	21	468,678,664.8	1.63	23,141,686.6	0.08
Library	8	131,063,388.6	0.46	14,788,482.9	0.05
Lifestyle center	1	4,753,803.1	0.02	1,573,937.3	0.01
Mall	8	80,657,950.2	0.28	26,641,660.2	0.09
Medical office	14	213,490,519.9	0.74	83,535,323.1	0.29
Mixed use property	30	965,741,600.0	3.36	461,484,334.6	1.62
Movie theater	3	22,801,218.9	0.08	10,514,599.8	0.04
Multifamily housing	1331	7,611,278,688.9	26.46	13,849,357,181.9	48.59
Museum	5	165,827,303.5	0.58	211,076,547.7	0.74
Office	340	8,944,429,927.9	31.09	2,635,616,230.5	9.25
Other	14	191,227,680.4	0.66	156,715,974.1	0.55
Performing arts	7	60,193,105.1	0.21	23,661,439.1	0.08
Physical therapy	2	11,318,342.3	0.04	8,709,848.2	0.03
Preschool/daycare	1	4,188,120.8	0.01	2,528,453.8	0.01
Prison	3	209,458,507.1	0.73	488,653,665.4	1.71
Public assembly	6	137,503,177.9	0.48	104,571,594.7	0.37
Public services	1	7,293,259.2	0.03	9,532,365.8	0.03
Recreation	16	42,309,927.6	0.15	134,934,393.5	0.47
Repair services	2	4,447,996.5	0.02	6,073,873.1	0.02
Residence hall	26	168,017,563.5	0.58	156,724,509.6	0.55
Residential	4	20,072,114.7	0.07	34,611,486.1	0.12
Residential care	11	51,846,514.1	0.18	83,011,165.1	0.29
Retail store	56	508,360,390.5	1.77	244,094,970.0	0.86
Senior care	53	166,765,835.1	0.58	306,709,397.2	1.08
Social/meeting hall	2	13,562,048.3	0.05	17,265,782.8	0.06
Specialty hospital	1	8,478,594.8	0.03	13,801,832.7	0.05
Strip mall	26	179,545,110.1	0.62	80,152,943.7	0.28
Supermarket/grocery	45	480,065,598.9	1.67	270,007,557.2	0.95
Wholesale club	8	95,055,846.2	0.33	64,954,953.2	0.23
Worship facility	14	34,068,836.7	0.12	56,200,787.4	0.20
Total percentages	—	—	100.00	—	100.00

Table 2. Distribution of zones of source EUI for hospitals, multifamily housing, office, and university buildings located within Chicago Metropolitan Area

Building type	No. of buildings	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
Hospital	30	0	0	0	0	30
Multifamily housing	1,331	334	335	299	244	119
Office	340	20	21	44	93	162
University/	87	3	2	14	21	47
College	1,788	357	358	357	358	358

Note: Analysis covering Chicago Metropolitan Area (55 zip codes).

Table 3. Distribution of zones of source EUI (source energy usage index) for hospitals, multifamily housing, office and university buildings located within zip code 60616

Building type	No. of buildings	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
Hospital	1	0	0	0	0	1
Multifamily housing	75	14	16	22	17	6
Office	3	0	2	0	0	1
University/	2	0	0	0	1	1
College	81	14	18	22	18	9

Note: Focused analysis on zip code 60616.

to a school. DataMade's 2nd City Zoning (DataMade 2019) specifies nine zoning classes for Chicago: (1) residential, (2) business, (3) commercial, (4) downtown, (5) manufacturing, (6) planned manufacturing, (7) planned development, (8) transportation, and (9) parks and open space. Each of these classes can be narrowed down to subclasses. As a case in point, the subcategories for residential are (1.1) detached, single-family homes (RS); (1.2) two-flats, townhouses, low-density apartment buildings, single-family homes

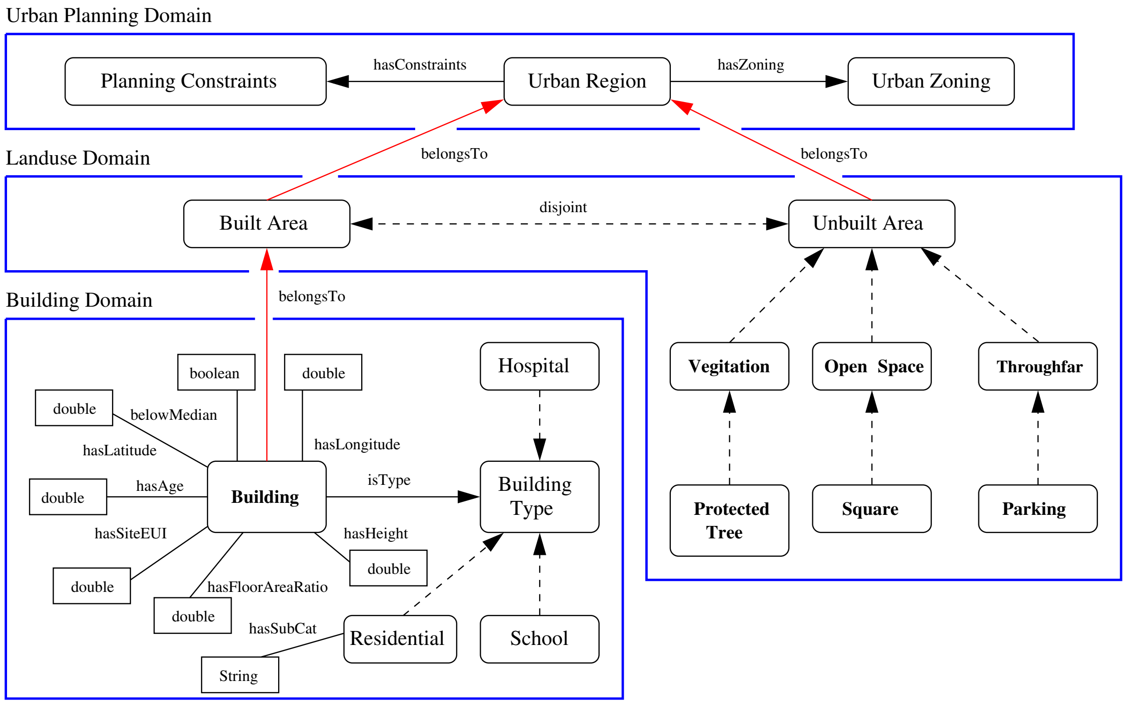

Fig. 7. Simplified semantic model for urban planning, land use, and building domains.

```
Rules and Associations for Zoning of Residential Buildings

Association 1: Site EUI (kBtu/sq ft)='(-inf-56.55]'
  ==> Building Type=Multifamily Housing   <conf:(0.84)>

Association 2: Community Area=NEAR NORTH SIDE
  ==> Building Type=Multifamily Housing <conf:(0.8)>

Association 3: Gross Floor Area - Buildings (sq ft)='[150354-513214]'
  ==> Building Type=Multifamily Housing <conf:(0.83)>

Association 4: Year Built='(1999.5-inf)'
  ==> Primary Property Type=Multifamily Housing <conf:(0.85)>

Rule 1:Building(?x), hasFloorAreaRatio(?x,?a), greaterThan(?a,6.6)
       isType(?x,?t), equal(?t,"multi-family") -> hasSubCat(?t,"RM6.5")

Rule 2:Building(?x) hasAge(?x,?a)
       greaterThan(?a,20) -> isType(?x,"multi-family")
```

Fig. 8. Urban planning semantic rules associations for residential units as a function of building attribute [EUI ($\mathrm{kBtu/ft^2} = 3.14\ \mathrm{kWh/m^2}$), gross floor area ($\mathrm{ft^2} = 0.092\ \mathrm{m^2}$), year built] values.

(RT); and (1.3) medium- to high-density apartment buildings (multifamily housing), two-flats, townhouses, and single-family homes (residential multiunit housing). The urban ontology includes all these classes, their subclasses, and rules to fill in semantic relationships based on characteristics of attributes in the data.

Fig. 8 highlights the strength of the proposed framework, where results of feature association mined from the data set work in tandem with the semantic domain to provide more information regarding the real world. Using an a priori type of algorithm, four associations between the building type and site EUI, community area, gross floor area, and building age were obtained at an 80% confidence level. To obtain further knowledge about the exact type, Rule 1 narrows down the type to "RM-6.5," a definition described in DataMade (2019). Data to compute floor area ratio, building height, and footprint will be obtained from OpenStreetMap. It is important to note that the association rules obtained from ML are consistent with Web-based zoning maps for the city of Chicago. Moreover, based on Association 4, Rule 2 identifies the building type based on age. One can reasonably expect that as new buildings are constructed in Chicago, the applicability of this association relationship will evolve over time. Thus, from both urban and city planning perspectives, the digital twin can be viewed as a tool that can determine the extent to which architectural data stored in online databases (e.g., OpenStreetMap) are consistent with current zoning models.

```
Experiment A
eui <= 171.5    <-- first breakpoint
|   zip = 60616
|   |   age <= 102: MULTIFAMILY HOUSING (59.0)
|   |   age > 102
|   |   |   eui <= 118.2: MULTIFAMILY HOUSING (5.0)
|   |   |   eui > 118.2: OFFICE (3.0/1.0)
eui > 171.5
|   zip = 60616
|   |   eui <= 269.2: MULTIFAMILY HOUSING (11.0/1.0)
|   |   eui > 269.2: COLLEGE/UNIVERSITY (3.0/2.0)

Number of Leaves: 138
Size of the tree: 153
Correctly Classified Instances  1443 --> 80.7047%
Incorrectly Classified Instances 345 --> 19.2953%
```

(a)

```
Experiment B
zip = 60616    <-- first breakpoint
|   age <= 86
|   |   age <= 53
|   |   |   area <= 115066: NEAR SOUTH SIDE (13.0/6.0)
|   |   |   area > 115066
|   |   |   |   age <= 12: NEAR SOUTH SIDE (5.0)
|   |   |   |   age > 12
|   |   |   |   |   eui <= 130.2: DOUGLAS (3.0)
|   |   |   |   |   eui > 130.2: NEAR SOUTH SIDE (8.0/2.0)
|   |   age > 53: DOUGLAS (18.0/2.0)
|   age > 86: NEAR SOUTH SIDE (7.0/2.0)

Number of Leaves:  82
Size of the tree: 102
Correctly Classified Instances  1399  --> 78.2438%
Incorrectly Classified Instances 389  --> 21.7562%
```

(b)

Fig. 9. Results of J48 classification and reduced error pruning for zip code 60616. Experiments A and B reveal features [EUI ($\text{kBtu/ft}^2 = 3.14\ \text{kWh/m}^2$), age (years), zip code] that (a) distinguish various types of buildings; and (b) best describe buildings within community areas.

Data Mining Experiments

Classification analyses are performed with the J48 algorithm (Kaur and Chhabra 2014) on the complete set of buildings covered in Table 2. The analysis begins with a divide-and-conquer strategy to organize the data into a multiattribute tree hierarchy, organized to maximize information gain at each successive layer of the tree structure. The second stage of the analysis involves simplification (i.e., application of a reduced error pruning algorithm) of the decision tree structure to obtain a desirable balance of accuracy versus compactness. A basic question is whether the Chicago Energy Benchmarking data set contains features that strongly contribute to building energy consumption. To investigate this issue, three experiments were conducted.

Experiment A

The first experiment focuses on the identification of attributes that distinguish the four types of buildings under consideration: hospitals, multifamily housing, offices, and universities and colleges. Potentially good attributes for classification include zip code, latitude and longitude, EUI value, floor area, and age of building. Fig. 9(a) shows a sample of the results, focusing on buildings that are located within zip code 60616. Here, the most important breakpoint is clearly EUI value, followed by zip code and lower-level combinations of building age and refinements to interval ranges for the EUI value. The classification of 1,788 buildings generates a binary decision tree that has 138 leaves and correctly classifies the building instances with 80% accuracy.

Experiment B

The second experiment [Fig. 9(b)] focuses on the identification of parameters that best describe buildings within community areas. Zip code 60616 is home to two community areas, Douglas and Near South Side. Building instances are classified with an accuracy of 78% using a decision tree that has 82 leaves spanning the 55 zip codes and lower-level parameters involving floor area, building age, and EUI value. This procedure provides insight into the nature of the buildings within a specific community area that is beyond a simple spreadsheet analysis. As a case in point, it is possible to now say that within 60616, the community area Douglas is defined by two categories of building: (1) 18 buildings constructed during the period 1933 to 1963 (i.e., $53 < \text{age} \leq 83$) and (2) 3 large buildings with low/medium energy usage, constructed during the 1953–2004 period (i.e., $115{,}066 < \text{area}\ \&\&\ 12 < \text{age} \leq 53\ \&\&\ \text{eui} \leq 130$). The results of this classification illustrate a situation where complex decision trees are pruned to remove redundant tests before transformation into rules.

Experiment C

The final experiment investigates opportunities for using attributes within the Chicago Energy Benchmarking data set (property type, zip code, age, floor area, latitude and longitude) as rules for the classification of EUI zones. Previous research (EnergyStar 2018) demonstrated significant variation in energy usage across building types. The contents of Tables 2 and 3 indicate, however, that significant variations in energy usage can also occur within a single type of building. Using property type and EUI zone of energy usage as target features for classification, Weka simply predicts EUI Zone 2. The number of correctly classified building instances are 353, with an accuracy level of 19.7% and in line with the distribution of building inventory shown in Table 2. Adding building age and floor area separately to the set of classification attributes increases the prediction accuracy to only 27% and 26%, respectively. Similar computational results occur for the analysis of office buildings and multifamily buildings alone. Collectively, these results indicate that floor area and building age are poor predictors of source EUI and, thus, are unlikely to find a role in the development of semantic rules governing the energy management of buildings.

Discussion

The case study analyzed the Chicago Energy Benchmarking data set from data mining (ML) and semantic modeling perspectives, with an eye to understanding the extent to which the data support the classification of buildings at the neighborhood (or community area or zip code) level and the prediction of building energy usage as a function of the data set attributes or geographical location. Knowledge of such trends, even if they are subject to uncertainty, are useful for stakeholders concerned with energy consumption (e.g., building managers, energy companies, city managers). These trends also allow designers of digital twin systems to understand how domain rules should be structured and to identify the details of interactions among the ML and semantic representations.

The case study procedure began with spreadsheet-like computations for high-level examination of the data set. This was followed by use of Apache Jena for the construction of preliminary semantic graph models and rules (covering building, land use, and urban domains), and Weka (Witten et al. 2017) for the data mining experiments. The key observations from the preliminary study are as follows. First, it is evident that real-world data sources introduce a multitude of complexities into the analysis, some of them unforeseen. For example, the original idea was that perhaps it would be possible to look for relationships between community area and various metrics of energy usage. But it turned out that community area boundaries were not precisely defined, and this necessitated removing "community areas" from consideration and working with zip code regions instead (Fig. 6). Second, it is well known that certain types of buildings consume more energy than others, but even within a single category of building (e.g., multifamily housing) there is considerable scatter in the energy usage data (Table 3). The magnitude of these variations highlights weaknesses in state-of-the-art classification procedures based on median values alone. They also complicate ML procedures aimed at resolving questions (e.g., does the Chicago Energy Benchmarking data set contain features that strongly contribute to building energy consumption?) one might have about the factors affecting energy usage in buildings.

Conclusions and Future Work

The authors' long-term vision for smart city digital twins centers around the development of a so-called city operating system that provides city stakeholders with enhanced levels of situational awareness and decision-making support for the management of urban infrastructure and services embedded in space-time terrain. In a step toward providing this capability, this paper proposed a digital twin architecture where knowledge representation and reasoning and ML formalisms work in tandem, providing complementary and supportive roles in the collection and processing of data. Both sides of the proposed architecture rely on streams of urban data/information corralled into data system hubs. Computations with a priori algorithms resulted in reasonably accurate (80% accuracy) characterizations of the types of buildings found in various zip codes along with their energy usage. Decision tree algorithms also identified with 78% accuracy the roles that building age and floor area play in describing buildings found in the various community areas. Both of these analyses provide valuable insight on the correct usage of parameters to efficiently characterize the buildings found in various neighborhoods. However, Weka failed to find a strong correlation between building energy usage and parameters such as building age, floor area, and location. This does not mean that such a relationship does not exist but rather points to a strong need for expansion of the data set and analyses to include factors (e.g., type of construction, building geometry) not covered by the Chicago Benchmark data set.

The development of urban data system hubs, a precursor to digital twins, is rapidly expanding among the world's cities. As a case in point, a data-driven city operations center in Rio de Janeiro, Brazil, pulls together into a single location real-time data streams from 30 agencies, including transportation, utility, emergency, and security services. These data are overseen by a staff of 180 data operatives, who monitor and react to urban and environmental processes. City operating systems that completely bypass humans to manage communication between systems are also being explored. Projects such as Microsoft's CityNext, IBM's Smarter City, Urbiotica's City Operating System, and PlanIT's Urban Operating System have been working toward this possibility (Kitchin 2015). Nevertheless, city operating systems face significant challenges in communication, security, availability, resiliency, energy efficiency, network bandwidth, focus on citizens, and levels of standardization (Mora and Larios 2015). Extracting meaningful information from urban system exchanges is complicated by a variety of factors, including heterogeneity of the data and the volume and rate of data generation. To address these challenges, future work needs to consider the capabilities of ML technologies that extend beyond those shown at the bottom of Figs. 4 and 5. This work should include the use of deep learning and node2vec representations to learn the structure of urban networks (Patterson and Gibson 2017). Significant opportunities also exist for improving urban planning procedures through the integration and processing of collected data that are relevant to decision-making procedures. For example, spatial distributions of "urban plan area density" can be computed from geographical attributes and various building metrics and properties (e.g., latitude, longitude, height, width), but such metrics are not commonly used in state-of-the-art urban planning procedures. Looking forward, a significant challenge exists in integrating the data in these different domains and making relevant deductions about what they imply.

Acknowledgments

The authors gratefully acknowledge the assistance of Andrew Persily, Steven Bushby, Amanda Pertzborn, and Jiying Liu who reviewed and commented on the manuscript draft. The authors acknowledge the National Research Council (NRC) Associates Program as a funding source for Parastoo Delgoshaei.

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National

Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

Alam, K. M., and A. El Saddik. 2017. "C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems." *IEEE Access* 5: 2050–2062. https://doi.org/10.1109/ACCESS.2017.2657006.

Apache Jena. 2016. "An open source java framework for building Semantic Web and linked data applications." Accessed March 1, 2019. https://jena.apache.org/.

Arp, R., B. Smith, and A. D. Spear. 2015. *Building ontologies with basic formal ontology.* Cambridge, MA: MIT Press.

Austin, M. A., P. Delgoshaei, and A. Nguyen. 2015. "Distributed systems behavior modeling with ontologies, rules and message passing mechanisms." In Vol. 44 of *Proc., 13th Annual Conf. on Systems Engineering Research (CSER 2015)*, 373–382. Hoboken, NJ: Elsevier.

Austin, M. A., V. Mayank, and N. Shmunis. 2006. "PaladinRM: Graph-based visualization of requirements organized for team-based design." *Syst. Eng.* 9 (2): 129–145. https://doi.org/10.1002/sys.20049.

Baader, F., I. Horrocks, and U. Sattler. 2008. "Description logics." Chap. 3 in *Handbook of knowledge representation*, edited by F. van Harmelen, V. Lifschitz, and B. Porter, 135–180. New York: Elsevier.

Bello, J. P., C. Silva, O. Nov, L. Dubois, A. Arora, J. Salamon, C. Mydlarz, and H. Doraiswamy. 2019. "SONYC: A system for monitoring, analyzing, and mitigating urban noise pollution." *Commun. ACM* 62 (2): 68–77. https://doi.org/10.1145/3224204.

Berners-Lee, T., J. Hendler, and O. Lassa. 2001. "The Semantic Web." *Sci. Am.* 284 (5): 34–43.

Boschert, S., R. Rosen, and G. Heinrich. 2018. "Next generation digital twin." In *Proc., 12th Int. Symp. on Tools and Methods of Competitive Engineering (TMCE 2018)*, edited by I. Horvath, J. P. Suarez Rivero, and P. M. Hernandez Castellano, 209–218. Las Palmas de Gran Canaria, Spain: Delft Univ. of Technology.

Cai, H., V. W. Zheng, and K. C. C. Chang. 2017. "A comprehensive survey of graph embedding: Problems, techniques and applications." *IEEE Trans. Knowl. Data Eng.* 30 (9): 1616–1637. https://doi.org/10.1109/TKDE.2018.2807452.

Casey, M. J., and M. A. Austin. 2002. "Semantic Web methodologies for spatial decision support." In *Proc., Int. Conf. on Decision Making and Decision Support in the Internet Age (DSI-Age 2002)*, 453–463. Cork, Ireland: Oaktree Press.

Chicago. 2019. "Chicago Energy Benchmarking dataset." Accessed December 17, 2019. https://www.chicago.gov/city/en/depts/mayor/supp_info/chicago-energy-benchmarking/Chicago_Energy_Benchmarking_Reports_Data.html.

Coelho, M., M. A. Austin, and M. R. Blackburn. 2017. "Semantic behavior modeling and event-driven reasoning for urban system of systems." *Int. J. Adv. Intell. Syst.* 10 (3 and 4): 365–382.

DataMade. 2019. "2nd city zoning." Accessed February 23, 2019. http://secondcityzoning.org.

Delgoshaei, P., M. Heidarinejad, and M. A. Austin. 2018. "Combined ontology-driven and machine learning approach to monitoring of building energy consumption." In *Proc., 2018 Building Performance Modeling Conf. and SimBuild*. Chicago: ASHRAE and IBPSA-USA.

EIA (Energy Information Administration). 2016. "Frequently asked questions (FAQS)." Accessed March 1, 2019. https://www.eia.gov/tools/faqs/.

EnergyStar. 2018. *US energy use intensity by property type*. Washington, DC: USEPA.

Falquet, G., C. Metral, J. Teller, and C. Tweed. 2005. *Ontologies in urban development projects*. London: Springer.

Glaessgen, E. H., and D. S. Stargel. 2012. "The digital twin paradigm for future NASA and US Air Force vehicles." In *Proc., 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conf.* New York: American Institute of Aeronautics and Astronautics.

Goldsmith, S., and S. Crawford. 2014. *The responsive city: Engaging communities through data-smart governance*. San Francisco: Jossey-Bass.

Hendler, J. 2001. "Agents and the Semantic Web." Accessed April 4, 2002. http://www.computer.org/intelligent.

Hong, T., H. Kim, and T. Kwak. 2012. "Energy-saving techniques for reducing CO_2 emissions in elementary schools." *J. Manage. Eng.* 28 (1): 39–50. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000073.

Jin, R., C. Hancock, L. Tang, C. Chen, D. Wanatowski, and L. Yang. 2017. "Empirical study of BIM implementation-based perceptions among Chinese practitioners." *J. Manage. Eng.* 33 (5): 04017025. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000538.

Kaewunruen, S., P. Rungskunroch, and J. Welsh. 2018. "A digital-twin evaluation of net zero energy building for existing buildings." *Sustainability* 11 (1): 159. https://doi.org/10.3390/su11010159.

Kaur, G., and A. Chhabra. 2014. "Improved J48 classification algorithm for the prediction of diabetes." *Int. J. Comput. Appl.* 98 (22): 13–17. https://doi.org/10.5120/17314-7433.

Kitchin, R. 2015. "Data-driven, networked urbanism." In *Proc., Data and the City Workshop*. Maynooth, Ireland: Maynooth Univ.

Lee, J., B. Bagheri, and H. Kao. 2015. "A cyber-physical systems architecture for industry 4.0-based manufacturing systems." *Manuf. Lett.* 3 (Jan): 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001.

M2M (Machine-to-Machine Communications). 2015. *Impact of smart city activity on IoT environment*. Sophia Antipolis, France: European Telecommunications Standards Institute.

Mohammadi, N., and J. E. Taylor. 2018. "Smart city digital twins." In *Proc., 2017 IEEE Symp. Series on Computational Intelligence (SSCI)*, 1–5. Honolulu: IEEE. https://doi.org/10.1109/SSCI.2017.8285439.

Mora, O. B., and V. M. Larios. 2015. "Urban operating systems for sensor network management in smart cities." In *Proc., IEEE Guadalajara GDL CCD White Papers*. Guadalajara, Mexico: IEEE.

Negri, E., L. Fumagalli, and M. Macchi. 2017. "A review of the roles of digital twin in CPS-based production systems." *Procedia Manuf.* 11: 939–948. https://doi.org/10.1016/j.promfg.2017.07.198.

Nie, J., W. S. Xu, D. Z. Cheng, and Y. L. You. 2019. "Digital twin-based smart building management and control framework." In *Proc., 2019 Int. Conf. on Artificial Intelligence and Computing Science (ICAICS 2019)*. Lancaster, PA: DEStech Transactions on Computer Science and Engineering.

OSM (Open Street Map). 2019. "Open street map." Accessed January 25, 2019. https://www.openstreetmap.org.

Patterson, J., and A. Gibson. 2017. *Deep learning: A practioner's approach*. Sebastopol, CA: O'Reilly Media.

Petnga, L., and M. A. Austin. 2016. "An ontological framework for knowledge modeling and decision support in cyber-physical systems." *Adv. Eng. Inf.* 30 (1): 77–94. https://doi.org/10.1016/j.aei.2015.12.003.

Preuss, P. 2017. *Smart moves: California's next-gen infrastructure*. Berkeley, CA: Univ. of California, College of Engineering.

Privat, G., T. Coupaye, S. Bolle, and P. Raipin-Parvedy. 2019. "WoT graph as multiscale digital-twin for cyber-physical systems-of-systems." In *Proc., 2nd W3C Web of Things Workshop*. Grenoble, France: Gilles Private Orange Labs Services.

Ramaswami, A., L. Bettencourt, A. Clarens, S. Das, G. Fitzgerald, E. Irwin, D. Pataki, S. Pinceti, K. Seto, and P. Waddell. 2018. *Sustainable urban systems: Articulating a long-term convergence research agenda*. A Report by the Advisory Committee for Environmental Research and Education. Washington, DC: National Science Foundation.

Rinaldi, S. M., J. M. Peerenboom, and T. K. Kelly. 2001. "Identifying, understanding, and analyzing critical infrastructure interdependencies." *IEEE Control Syst. Mag.* 21 (6): 11–25.

Savisalo, A., J. Salmelin, J. Hemila, and K. Tuukkanen. 2018. "Digital twin as a city management tool." In *Proc., World Bank Conf. on Land and Poverty*, Washington DC: World Bank.

Selberg, S., and M. A. Austin. 2008. "Toward an evolutionary system of systems architecture." In Vol. 18 of *Proc., INCOSE Int. Symp., International Council of Systems Engineering*, 1065–1078. Utrecht, Netherlands: Wiley.

Shetty, S. 2017. "How to use digital twins in your IoT strategy." Accessed February 10, 2019. http://www.gartner.com/smarterwithgartner/how-to-use-digital-twins-in-your-iot-strategy.

Shrestha, P., and P. Kulkarni. 2013. "Factors influencing energy consumption of energy star and non-energy star homes." *J. Manage. Eng.* 29 (3): 269–278. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000134.

Shrivastava, S., and A. Chini. 2016. "Contractor energy information system: Energy-savings tool for building construction." *J. Manage. Eng.* 32 (3): 04015048. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000412.

Tahmasebinia, F., D. Fogerty, L. O. Wu, Z. Li, S. M. E. Sepasgozar, K. Zhang, S. Sepasgozar, and F. A. Marroquin. 2019. "Numerical analysis of the creep and shrinkage experienced in the Sydney Opera House and the rise of digital twin as future monitoring technology." *Buildings* 9 (6): 137. https://doi.org/10.3390/buildings9060137.

Tao, F., J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui. 2018. "Digital twin-driven product design, manufacturing and service with big data." *Int. J. Adv. Manuf. Technol.* 94 (9): 3563–3576. https://doi.org/10.1007/s00170-017-0233-1.

Taylor, J. E., and P. G. Bernstein. 2009. "Paradigm trajectories of building information modeling practice in project networks." *J. Manage. Eng.* 25 (2): 69–76. https://doi.org/10.1061/(ASCE)0742-597X(2009)25:2(69).

Witten, I. H., E. Frank, M. A. Hall, and J. P. Christopher. 2017. *Data mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kauffmann.

Zheng, A., and A. Casari. 2018. *Feature engineering for machine learning*. Sebastopol, CA: O'Reilly Media.