Springer Optimization and Its Applications 186

Panos M. Pardalos
Stamatina Th. Rassia
Arsenios Tsokas *Editors*

Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities

Designing for Sustainability

Springer

Springer Optimization and Its Applications

Volume 186

Aims and Scope

Optimization has continued to expand in all directions at an astonishing rate. New algorithmic and theoretical techniques are continually developing and the diffusion into other disciplines is proceeding at a rapid pace, with a spot light on machine learning, artificial intelligence, and quantum computing. Our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in areas not limited to applied mathematics, engineering, medicine, economics, computer science, operations research, and other sciences.

The series **Springer Optimization and Its Applications (SOIA)** aims to publish state-of-the-art expository works (monographs, contributed volumes, textbooks, handbooks) that focus on theory, methods, and applications of optimization. Topics covered include, but are not limited to, nonlinear optimization, combinatorial optimization, continuous optimization, stochastic optimization, Bayesian optimization, optimal control, discrete optimization, multi-objective optimization, and more. New to the series portfolio include Works at the intersection of optimization and machine learning, artificial intelligence, and quantum computing.

Volumes from this series are indexed by Web of Science, zbMATH, Mathematical Reviews, and SCOPUS.

More information about this series at http://www.springer.com/series/7393

Panos M. Pardalos • Stamatina Th. Rassia
Arsenios Tsokas
Editors

Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities

Designing for Sustainability

Springer

Editors
Panos M. Pardalos
Department of Industrial and Systems Engineering
University of Florida
Gainesville, FL, USA

Stamatina Th. Rassia
University of Macedonia &
Le.D.R.A. Group
Greece

Arsenios Tsokas
Center for Applied Optimization
University of Florida
Gainesville, FL, USA

ISSN 1931-6828 ISSN 1931-6836 (electronic)
Springer Optimization and Its Applications
ISBN 978-3-030-84458-5 ISBN 978-3-030-84459-2 (eBook)
https://doi.org/10.1007/978-3-030-84459-2

Mathematics Subject Classification: 00-01, 62Hxx, 62-07, 62-04, 94Axx

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities: Designing for Sustainability presents a unique interdisciplinary combination of architectural, engineering, physics, mathematical, computer science, and other fields of strategic planning, marketing, and more. Its aim is to create a platform for the exchange of ideas on its broad and fascinating subject. Artificial intelligence, machine learning, and their respective optimization tools are created to offer an overarching smartness to the built environment that is yet to be converted into ever more sustainable. While each and every part of its title forms a topic of scientific endeavor by itself, the editors of this book considered combining them with the aim to identify a few interdisciplinary lines of expert thinking that can be useful to a wider audience working on these and related research pathways.

The chapters composing this book are written by eminent researchers and practitioners who present their research results and ideas based on their expertise. Thus, in this book, a wide spectrum of topics is presented that appear under the following titles, written by the respective authors and author groups.

- **Cities as Convergent Autopoietic Systems**, *by Christopher G. Kirwan and Stefan V. Dobrev*
- **Digital 'Vitalism' and its 'Epistemic' Predecessors: 'Smart' Neoteric History and Contemporary Approaches**, *by Konstantinos Moraitis*
- **Unbuildable Cities***, by Stamatina Th. Rassia*
- **Smart Cities as Identities***, by Nikolaos Tsoniotis*
- **A Cross-Domain Landscape of ICT Services in Smart Cities***, by Barbora Bunhova, Terezia Kazickova, Mouzhi Ge, Leonard Waletzky, Francesco Caputo, and Luca Carrubbo*
- **A Novel Data Representation Method for Smart Cities' Big Data***, by Attila N. Nagy and Vilmos Simon*
- **A Pedestrian Level Strategy to Minimize Outdoor Sunlight Exposure***, by Xiaojiang Li, Yuji Yoshimura, Wei Tu, and Carlo Ratti*
- **Planning and Management of Charging Facilities for Electric Vehicle Sharing***, by Long He, Guangrui Ma, Wei Qi, Xin Wang, and Shuaikun Hou*

- **A Reactive Architectural Proposal for Fog/Edge Computing in the Internet of Things Paradigm with Application in Deep Learning**, *by Oscar Belmonte-Fernandez, Emilio Sansano-Sansano, Sergio Trilles, and Antonio Caballer-Miedes*
- **Urban Big Data: City Management and Real Estate Markets**, *by Richard Barkham, Sheharyar Bokhari, and Albert Saiz*
- **Social Media-Based Intelligence for Disaster Response and Management in Smart Cities**, *by Shaheen Khatoon, Amna Asif, Md Maruf Hasan, and Majed Alshamari*

We would like to express our special thanks to all the authors of the chapters contributed in this book. Last but not least, we would like to acknowledge the superb assistance of the Springer staff during the preparation of this publication.

Gainesville, Florida, USA — Panos M. Pardalos

Greece — Stamatina Th. Rassia

Gainesville, Florida, USA — Arsenios Tsokas

Contents

Cities as Convergent Autopoietic Systems

Christopher G. Kirwan and Stefan V. Dobrev

1 Introduction

The evolution of the human-constructed environment is a roadmap to understanding the impact of technology on nature and the resulting symbiotic relationship of humans, cities, and the natural world. Prior to the industrial revolution, human habitation relied on the delicate balance with nature, and as such, most concentrations of the human population lived in harmony with the environment from which it sustained essential resources. The nineteenth century was the century of industrialization that allowed the vast exploitation of physical resources and the construction of extensive infrastructure across the world. This new expansion disrupted the delicate balance between humans and nature and gave rise to unsustainable resource-intensive industrial development that propelled the twentieth century's explosive growth of the modern urban metropolis. The twenty-first century will be defined by carbon-neutral innovation, and the design and development of sustainable smart cities. If managed correctly on both global and local levels, new intelligent urban systems will allow humans and their habitat to return to a more harmonious relationship.

Our renewed symbiosis with nature will drive the transformation of cities in the twenty-first century, reaffirming the critical role cities must play in the fight against climate change and planetary degradation. Cities offer many competitive advantages over nations when it comes to responding to current and emerging challenges—flexibility, adaptability, access to capital, leading technology, educated workforce,

C. G. Kirwan
Henley Business School, University of Reading, Informatics Research Centre, Reading, UK
e-mail: cgk@kiirwandesigm.com

S. V. Dobrev (✉)
Maastricht University, School of Business and Economics, London, UK

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_1

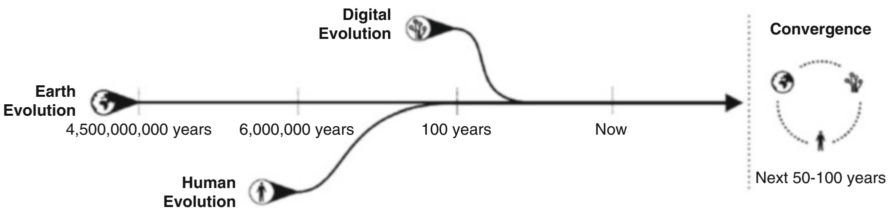


Fig. 1 Convergence evolution—earth, human, digital (Adapted from Kirwan and Fu 2020)

transportation, and communications infrastructure. Furthermore, cities are strategically positioned to benefit from the Fourth Industrial Revolution, where humans and machines will become increasingly connected, reaching an unprecedented level in the form of advanced co-design and co-development to accelerate innovation, commonwealth, quality of life, and individual well-being. To achieve the highest potential outcomes for humanity and planet earth, this next stage in humanity's evolution will require a deeper understanding of the nature of living systems and the potential application of convergent properties that guide the behavior of self-sustaining systems.

Autopoiesis, or the process of self-production and self-organization, is a theory introduced in the second part of the twentieth century to describe the characteristics and behaviors of living entities and biological systems. Humberto Maturana and Francisco Varela introduced the term and defined an autopoietic system as "*a network of inter-related component-producing processes such that the components in interaction generate the same network that produced them.*" (Geyer 1995, pg. 12). The concept has seen wide application in the fields of mathematics (Robert Rosen), the study of cognition (Maturana and Varela; Luhman; Mingers), and in studies of the nervous system, information systems, sociology (Luhman; Mingers; Luisi and Damiano), legal studies (Hilpold), biomimetics, and others.

Contemporary cities and their core functions are rapidly transforming as environmental, human, and technological systems converge (Fig. 1). In evolutionary biology, convergence describes the process by which different species of animals and plants develop similar features and characteristics, without sharing common ancestors, but as a response to environmental conditions and stimuli. This form of structural coupling of a living system or entity to its environment is a critical evolutionary mechanism for a system's survival. The tendency of a living system to alter its physical and organizational structure through perpetual interaction and exchange with its environment is a key trait of autopoietic systems, allowing living systems to maintain, self-organize, self-replicate, and adapt themselves to their environment by regulating their composition and preserving their boundary (Maturana and Varela 1980).

Through the autopoietic lens, cities can be visualized as complex living organisms composed of multiple functions working together as real-time operating systems. Building on the metaphor of the human body, it is possible to understand cities as complex gestalt formations of co-dependent subsystems (i.e. circulation, energy, waste management), individually functioning within the whole. Smart cities flow with real-time data collected from a variety of sources, including human, machine, and environmental input, each providing critical operating intelligence. Big data, representing the multiple functions of the body, augmented via artificial intelligence (AI), can be collected, filtered, and analyzed through feedback loops allowing city operations to perform as self-regulating ecosystems. The challenge has been linking these diverse subsystems within the city as a total integrated operating system since each of today's urban functions require different types of response mechanisms and processing from more human-centric to autonomous machine-driven intelligence.

To reinforce the comparison of city operating systems with systems in nature, we introduce two principal concepts to understand cities as autopoietic systems. The first concept is the city as a real-time flow of data in the form of an urban metabolism. In this representation, the city manages multiple functions simultaneously in real-time, responding to internal and external stimuli and conditions requiring an orchestrated reaction. The second concept is the city as a form of biomimicry. Biomimetic urban systems will develop behaviors that emulate nature to become more efficient, self-regulating systems. As technology becomes embedded ubiquitously in the urban fabric, cities and nature will become more interconnected and share common characteristics, including sensing, adapting, and responding to environmental conditions. This combination of urban metabolism and biomimetic urban systems forms the basis of the autopoietic city operating system.

The incorporation of autopoietic principles within the meta-architecture of Smart City operating systems (OS) will enable the transformation of cities and their complex functions to develop as self-regulating systems reminiscent of living organisms. We present our investigation across three dimensions established for the purpose of this paper—*Theory, Characteristics and Applications*. In the first section *Theory*, we present a critical review of academic literature and models relating to autopoiesis in the context of leading "System Theories." In the second section of the paper, *Characteristics,* we present a selection of key properties and behaviors of autopoiesis that we have extracted from the literature review. In the final section, *Applications*, we explore how these characteristics and behaviors will converge to accelerate the evolution of core smart city functions and operating systems we refer to as Autopoietic Operating Systems (AOS), a term we have introduced.

2 Theory: Evolution of Living Systems Thinking

In this section of the paper, we present a brief overview of the evolution of Systems Thinking presenting a selection of the most impactful works for the purpose of this paper, including *General Systems Theory, Cybernetics, Autopoiesis, Second-order Cybernetics,* and *Anticipatory Systems* drawing from the different classes of systems to synthesize the most important principles and characteristics that govern living systems. We start our research by introducing the theory of Cybernetics (Norbert

Wiener), Second-order Cybernetics and General Systems Theory (Bertalanffy). Key concepts and characteristics serve as the foundation of Maturana and Varela's Autopoiesis Theory—our research's central theme. The self-preserving nature of autopoietic systems is the result of structural determinism and structural coupling—underlining behavioral traits of all living systems. Autopoiesis theory finds varied and increasingly useful application beyond the domain of Biology, including in Sociology (Luhmann 2012; Mingers 1991; Damiano and Luisi 2010), Governance (Andrew Dunsire 1996), Law and Human Rights (Peter Hilpold 2011), Smart Cities (Kirwan and Fu 2020), and Biomimetics (Robert Rosen 1978, 1985). In Sect. 3 we present and briefly discuss their implications. In Sect. 4, we discuss how these characteristics converge across the six core smart city functions to form the basis of Autopoietic Operating Systems (AOS).

2.1 Cybernetics

Norbert Wiener popularized the concept of *Cybernetics* in the 1940s while researching the application of control theory in relation to complex living and non-living systems. Wiener defined *Cybernetics* as "*the scientific study of control and communication in the animal and the machine*" (Wiener 1948). The term "Cybernetics" is derived from the Greek language and translated in English as "the art of steering." Indeed, it offers a powerful framework for analyzing the properties and understanding the behavior of living systems. Wiener discovers that both living and non-living systems operate according to cybernetic principles—they require communication to achieve effective action through continuous internal and external feedback. In biological terms, the process of feedback takes place to maintain homeostasis, an equilibrium/optimal state of a system. This self-correcting mechanism is critical to survival because it drives adaptation to random environmental events and conditions. Their ability to change through corrective action and adaptation is achieved through a perpetual cycle of sensing, gathering information through a series of feedback loops, and comparing to the system's original goals, prior to undertaking corrective action in a continuous pursuit of homeostasis.

Second-order cybernetics refers to systems classified as entities that encapsulate the capacity to project their operations on the environment and on themselves, regardless of whether the system is represented by a group or an individual. These operations give birth to variety within the environment or within systems themselves. This aspect can be regarded as a consequence of systematic variation, rendering systems as recursive. In recursive systems, communications can be conveyed, and observations can be noticed. The differences that exist between first-order and second-order cybernetics have been examined by von Foerster and others including Pask, Varela, Umpleby, and Parsons. These dissimilarities highlight the relationship between the aim of a model and the goal of the modeler, the connection between systems that are autonomous and controlled systems, identifying links between variables within a system and the interaction between the observed system and the observer, and can be applied to various theories that embody social systems and hypotheses concerning the interaction between society and ideas. The latter relationship illustrates a difference that appears to illuminate Parsons' approach as a theorist concerned with first-order systems as well as the stability and the maintenance of systems. On the other hand, Luhman, as a cybernetician, was more interested in morphogenesis and change in second-order systems. (Geyer 1995).

The relationship between first- and second-order cybernetics defines a progression in systems organizational behavior from a linear command and control model to an organic and autonomous system that incorporates the observer. In this more holistic formation, the observer becomes part of the system itself, and hence part of its evolutionary trajectory. This recursive interaction between the observer and the system is an example of more complex and intelligent systems.

2.2 General Systems Theory

Von Bertalanffy introduced General System Theory in 1956. Building on Cybernetic theory, Von Bertalanffy defines a system as "*a complex of interacting elements*." He also contemplated the idea of thinking systems across all disciplines to discover broad principles that are valid in connection with all systems. The concept of "system" was introduced as a new scientific paradigm (which characterized classical science) and was related to the contrast between the mechanical and analytical paradigm. A notion of paramount importance concerning the general systems theory is the focus it places on interactions. The center within relationships indicates that a single autonomous element's behavior is unlike its behavior when other elements engage in interaction with the aforementioned element. The differentiation between closed, open, and isolated systems represents another fundamental principle. Within systems that are open, exchanges of matter, energy, information and people occur with the external environment. Within closed systems, the only exchanges that take place are those which involve energy. Systems that are isolated are characterized by the complete lack of exchanges of elements. Based on these fundamental concepts, diverse approaches began to develop as a result of the emergence of the general systems theory. These include open system theory, the viable system approaches, and the viable system models. The open system theory (OST) is concerned with the examination of the relationships between different organizations and the environments that they are a part of (Mele et al. 2010).

Bertalanffy's approach to analyzing complex systems emphasizes key concepts such as embeddedness within other larger systems, dynamic processes of self-organization, growth, and adaptation. Bertalanffy adopts a holistic approach to his analysis of living systems in stark contrast to the conventional and widely accepted *reductionist* view on complex phenomena prevalent during the earlier parts of the twentieth century. The reductionist philosophical stance attempts to interpret complex systems in a *gestalt* state as the sum of all parts or components, while General Systems Theory looks at complex systems holistically, whereby the whole is bigger than the sum of its parts. Systems that learn and adapt must engage successfully with their environments to maintain growth and their ability to adapt. Within such a dynamic relationship, certain systems exist for the sole purpose of supporting the effective functioning of other systems, thereby preventing their failure. Bertalanffy's work is widely recognized for its universality and application beyond its original focus on theoretical biology and cybernetics, to also include fields as diverse as sociology, economics, statistical analysis, ecology, meteorology, political science and psychology. Systems theory allows us to apply a common framework for the analysis and holistic understanding of complex phenomena and systems. As such, systems theory enables us to better understand individual components and subsystems in the context of their relationship to each other, as well as to other systems and their environment as a higher scale complex system.

2.3 The Theory of Autopoiesis

The term *autopoiesis* is derived from the Greek words *auto*, meaning "self" *and poiesis, meaning* "creation." The theory of autopoiesis was proposed by biologists Humberto Maturana and Francisco Varela in 1972 in their publication *Autopoiesis and Cognition: The Realization of the Living*, where they introduce their theory to describe the essential processes and characteristics that are fundamental for all living organisms. Autopoietic systems consist of self-creating processes that produce all components and subcomponents necessary to sustain its existence as a living entity.

Networks represent relationships between various components which are self-referential and generate the complexity that characterizes living organisms. These types of processes serve an identical function in the human body (i.e., a cell) as in the mind (i.e., cognition). Different components are incorporated in networks which are self-creating since they produce other components to sustain themselves as well as the structure in its whole complexity. In contrast, a system that harnesses the energy to generate complexity that is uncharacteristic of the system itself, is called "allopoiesis" —an example in that respect would be a factory assembly line whose elements manufacture external products, not to perpetuate itself. Varela and Maturana assert that the organism's primary function could be described as part of the nature of the self-referential as well as being encapsulated in the self-creating networks and processes. Pursuing a high-level understanding of neural networks, Maturana proposes that a similar process applies (Maturana and Varela 1980). Maturana ascertain that cognition represents a system governed by self-referencing, whereby understanding is shaped by previous understanding. What is grasped by the human mind through what the eyes perceive does not constitute the all-inclusive reality of the "outside"; rather, it is the mere articulation of the brain's neural networks that brings to light the experience of understanding and interpreting (Geyer 2001).

It is perhaps best understood in contrast to an allopoietic system, such as a factory, which takes in materials and uses them to produce something other than itself. Damiano and Luisi (2010), p.148 argue that artificial or completely abstract systems can also exhibit autopoiesis, however, to be considered "living, or alive" they must have cognitive capabilities.

The authors summarize three critical conditions and dimensions of autopoiesis as:

1. REACTION NETWORK: Self-production, self-organization, and self-maintenance. These properties are exhibited through a "regenerative network of processes which takes place within a boundary of its own making and regenerates itself through cognitive or adaptive interactions with the medium."
2. BOUNDARY: Defined by a semi-permeable boundary which allows for exchange with the system's external environment. The components of the boundary are being produced by a network of reactions which takes place within the boundary. The network of reactions is generated by conditions produced by the existence of the boundary itself.
3. COGNITION: The adaptive interaction of a living system with its environment.

2.3.1 Cognition

In their research on living systems, Maturana and Varela deal extensively with the topic of cognition. A central theme of autopoiesis theory is that of the continuous, recursive interaction of living systems with their surroundings. The authors define cognition as the "*recursive interaction of the autopoietic unit with its ambience. They described it as a self-regulating coupling of the system with the environmental context: an active coordination of the internal autopoietic processes with the environmental dynamics which allows the system to conservatively react to external variations. For Maturana and Varela it corresponds to a permanent self-determined modification of the system's patterns of activity which, seen from the outside, appears as «intelligent adaptive behavior» (selective assimilation of nutrients available in the environment, overcoming of obstacles while moving in the environment, etc.). Conceived this way, the act of cognition is an act which, literally, permits life, as it dynamically integrates the organic structure of the living systems in the environment. Maturana and Varela thus came to the strong conclusion that «life is cognition»—that there is no life without cognition.*" (Damiano and Luisi 2010, p. 148).

2.3.2 Structural Determinism

Structural determinism is the notion that form follows function. The structure of a system will determine its behavior, actions and evolutionary development. Whether we are exploring biological, artificial or abstract systems (e.g., legal, religious), structure (form) follows function. Systems adapt their form through the continuous pursuit of specific functions. Any changes that can occur to the structure of an autopoietic system must never interfere with the process of autopoiesis (self-regulation and self-reproduction), or it simply would not exist. Environmental forces could only “trigger or select” possibilities that the system’s structure makes available at any given time. The structure (the actual components and their relationships) may change dramatically over time or may be realized in many ways so long as the organization maintains its process of self-production. It can be said to be organizationally closed but structurally open (Mingers 1991, p. 280).

2.3.3 Structural Coupling

Structural coupling is the idea that autopoietic systems can become structurally coupled to other systems and their environments through a process that Maturana calls “evolutionary drift” or “mutual specification.” Since the system produces itself, it gains a significant degree of autonomy—it depends less on other entities for its continual existence. Simultaneously, if it ever fails to produce that which is necessary, then autopoiesis must break down, and the entity will disintegrate. If no functionalism is involved, however, the system either continuously maintains autopoiesis or does not (Mingers 1991, p. 280).

Sociologist Niklas Luhmann borrows the concept of autopoiesis and uses it to describe society as a complex autopoietic system. The social system, according to Luhmann, consists of communications which produce subsequent communications based on existing social structures. For instance, social conventions structure how to respond to questions, orders and statements. Such conventions, however, can never exhaustively determine the subsequent communications so that there is always an element of contingency involved in the system. That is, the system’s complexity entails that it must select some communications over others, i.e., the system must necessarily reduce complexity. The first and foremost way in which the social system reduces complexity is by drawing a boundary or distinction between itself and its outside, between the system and its environment. After that, the system can draw distinctions (i.e., communicate) on the system side of that distinction to increase its internal complexity. In this way, the social system is autopoietic. It generates its own elements as well as its own boundaries. Luhmann defines social systems not as a set of actions and functions but as the sum of all the information exchange taking place between all systems and subsystems (individual functioning systems such as the *economy, science, politics, media,* etc.). According to Luhmann’s view, all subsystems emerge from social systems, and as functions become increasingly differentiated, they achieve their own operational closure and autopoiesis. John Mingers (1991) examines the three elements of autopoietic systems (structural coupling, structural determinism, and boundary) within a social organizational perspective to determine the extent to which the original criteria and definitions of Maturana and Varela can be observed within a social context.

Mingers finds evidence that not only physical and biological, but also non-physical, artificial, and abstract systems can behave according to the principles of autopoiesis, as originally defined by Maturana and Varela. However, Mingers points out the necessity for a more abstract level of thinking when considering the systems' dimension, components, and boundary.

2.4 Anticipatory Systems

The theory of anticipatory systems was developed by mathematical biologist Robert Rosen in his book "*Anticipatory Systems: Philosophical, Mathematical & Methodological Foundations"* (1985). This book established the systematic study of the concept of "*anticipation"* in living systems and entities. According to Rosen, "*an anticipatory system is a natural system that contains an internal predictive model of itself and of its environment, which allows it to change state at an instant in accord with the model's predictions pertaining to a later instant. Note, in contrast, that a reactive system can only react, in the present, to changes that have already occurred in the causal chain, while an anticipatory system's present behavior involves aspects of past, present, and future. The presence of a predictive model serves precisely to pull the future into the present; a system with a 'good' model thus behaves in many ways as if it can anticipate the future. Model-based behavior requires an entirely new paradigm, an 'anticipatory paradigm', to accommodate it. This paradigm extends – but does not replace – the 'reactive paradigm' which has hitherto dominated the study of natural systems. The 'anticipatory paradigm' allows us a glimpse of new and important aspects of system behavior."* (Louie 2010, p. 20).

Prior to Rosen's theory, the general scholarly consensus was that natural systems ought to be regarded as "*reactionary"* rather than "*anticipatory"* in their nature and behavior. According to Louie (2010), the concept of anticipation has long been rejected since it violates the principles of causality, cause, and effect. In the context of Anticipatory Systems, it is worth mentioning the contributions of Edward Lorenz for the establishment of *chaos theory,* and more specifically, the deterministic nature of chaos. Lorenz is perhaps most famous for his metaphor of chaos theory, known as the "*butterfly effect."* This is the idea that within hyper-complex systems, the most minor action could drive a reaction within the wider system that is much greater in its magnitude and impact than could be anticipated. Chaos theory has revolutionized not only our understanding of complex natural systems, but also our ability to accurately predict their behavioral patterns, enabling humanity to better deal with complexity and manage systemic risks.

This brief literature review presents the evolution of living and intelligent system theory. Autopoiesis is a remarkable concept that can be observed in every living entity, as well as in completely abstract contexts such as social behavior and organizational structure. Our objective is to present a diverse range of concepts surrounding autopoiesis and convergence adapted to the scale and complexity of urban systems. In the next section, we curate key characteristics and properties of autopoiesis that form a meta-convergence as a new approach to the design and development of smart city operating systems.

3 Characteristics: Meta-Convergence

In Biology, the term evolutionary convergence is used to describe the process by which distinctly different animals develop similar traits and functional characteristics over time, without these traits being present in their last common ancestor. By exploring cities as living, autopoietic systems, we can introduce a new thinking/approach to urban functions, development and user experience that is based on a new convergent design and planning methodology grounded in the core theories presented in the previous section. The evolution of traditional cities to smart cities is a natural evolutionary progression driven by rapid advancements in our technological capabilities, accelerated by the increased global awareness and urgent mandate to reverse the challenges and effects due to climate change.

Having identified and presented several of the critical theories of systems organization and behavior in the previous section, we now turn our focus to defining the specific characteristics exhibited by autopoietic systems to determine to what extent such characteristics can be incorporated in the design, planning, and operations of future smart cities. Building on the essential attributes of autopoietic systems defined by Maturana and Varela, we propose the addition of several autopoietic properties. The convergence of these properties will enable the complexity and scalability we envision smart cities will require to manage diverse urban functions and dynamic urban growth as part of a gestalt operating system.

Table 1 represents the five characteristics and properties we have identified as key enablers of smart cities and the core attributes of autopoietic smart city operating systems. These properties work together to form a convergent composite we define as meta-convergence.

Table 1 Characteristics and properties of autopoietic systems

Characteristics and properties of autopoietic systems				
Sentience and cognition	Ambient intelligence	Anticipatory capabilities	Structural coupling	Structural determinism
Understanding through experiences and sensing Acquiring knowledge and expanding capabilities	Augmented natural environments Interactive, responsive and anticipatory Ubiquitous, non-physical based intelligence	Predictive capabilities Leading not reacting Agility and resilience Managing systemic risk and complexity	Interrelationship of objects to environment Symbiosis and codependency Adaptation through synergy	Form follows function Perpetual evolutionary mechanisms Continuous perfection & pursuit of optimal states

3.1 Sentience and Cognition

Maturana and Varela treat sentience and cognition as inseparable elements of how autopoietic systems operate. Such systems are structurally coupled with their environment and the ability to maintain recursive interactions with their environment, defined as medium, is key to their survival. The system function remains constant, while the structure adapts to their environment to maintain self-reproduction, organization, and reorganization. Through continuous recurring interactions, autopoietic systems gain knowledge. This is a key element to Maturana and Varela's autopoiesis theory—from the most basic organism to the most complex system, the process of cognition is a key property of living systems.

Perception and cognition occur through the operation of the nervous system, which is realized through the autopoiesis of the organism. As we have seen, autopoietic systems operate in a medium to which they are structurally coupled. Their survival is dependent on certain recurrent interactions continuing. For Maturana, this itself means that the organism has knowledge, even if only implicitly. The notion of cognition is extended to cover all the effective interactions that an organism has. A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. Living systems are cognitive systems and living as a process is a process of cognition. This statement is valid for all organisms, with and without a nervous system (Maturana and Varela 1980, p. 13).

As defined above, sentience and cognition are key characteristics of intelligent living systems. Through the rapid advancements in technology, machines are developing biomimetic capabilities that increasingly simulate behaviors observed in humans and other natural organisms. These capabilities will converge as a form of *collective intelligence* (Fig. 2) combining both natural and artificial characteristics propagating a new stage in the evolution of the relationship and communication of humans, machines and nature. As a result of these relationships, new communication capabilities and hybrid languages will form to support diverse combinations of human-machine-nature sentience and cognition.

3.2 Structural Coupling

As different components and systems become increasingly interconnected, new possibilities for training machine learning algorithms become available, giving rise to convergence driven by artificial intelligence. With advancements in technological innovation and increasingly technological convergence, coupled with the adoption of data-driven methodologies, systems are becoming increasingly smarter and autonomous. In the process of achieving convergence in systems, the underlying dynamics are the inherent forces in nature and evolution itself to

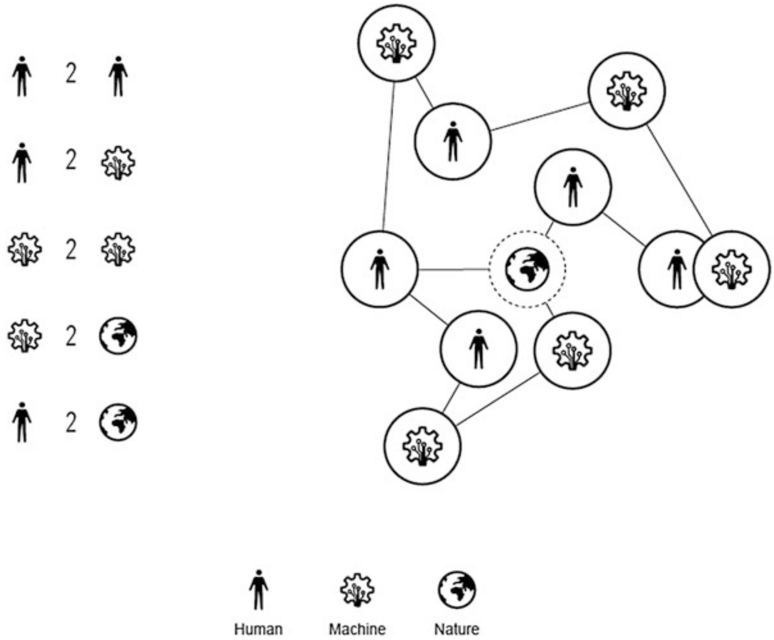


Fig. 2 Collective Intelligence (Adapted from Kirwan and Fu 2020)

develop parallel traits in systems and system behaviors. The diverse functions of systems and subsystems within smart cities, in principle, should move towards converging states given the same level of technological advancement and enablers, infrastructure/platforms and operational mechanisms. For example, if all smart cities' functions are developed on top of big data analytics, the behavior response mechanisms should, over time, develop convergence characteristics, including real-time data response, data filtering and processing. Applied to the difference between human and machine information processing, the convergence characteristics of autonomous (AI) data analytics will eventually lessen the disparities between human and machine data processing. Within this integrated domain, human and machine intelligence will be co-dependent and co-creative, allowing the generation of convergent responses across all smart city functions. Equally, with other AI-driven functions, human and machine intelligence will converge with the formation of a new hybrid intelligence that we have defined as *collective intelligence*. In their book on Smart Cities and AI, Kirwan and Fu (2020) expand the definition of collective intelligence to include a human-machine-nature convergence, ultimately integrating human and machine intelligence as part of a broader unified operating system that is the extension of the natural environment. This stage of urban evolution will enable autopoiesis to occur as a comprehensive, unified ecosystem, where smart city functions are harmoniously aligned to their physical environment while further expanded to interface with broader interplanetary systems.

3.3 Ambient Intelligence

In computing, ambient intelligence refers to a virtual environment that is sensitive and responsive to the presence of people. In smart cities, this refers to the development of ubiquitous sensor networks, or of a mesh of interconnected sensors, devices and technology *that embed input, processing, or response ubiquitously through the environment* Dainow (2017). Such interconnected environments will allow people to engage with their cities through seamless experiences rather than interacting with a disconnected set of individual components and discrete processes. These types of intelligent, interactive and adaptable environments are made possible through the convergence of domains and technologies, including IoT, Cloud Computing, machine learning, deep learning and artificial intelligence. What type of characteristics would ambient intelligence systems have? First, such systems will be ubiquitous and embedded, allowing users to engage, interact, and control their environments from any place and any device with a network connection. Second, these systems will be intelligent and clearly aware of the specific context, understanding exactly what is needed at any point in time, seeking ways to support users with information (warnings, updates, calculations, statistics, etc.), through ambient (light, temperature, humidity, etc.) and personalized recommendations. Such systems will be highly customizable to the individual user's interests, needs, and goals. Third, these systems must be adaptive, reactive and responsive. The ability of such systems to operate in real-time is a critical condition for their successful implementation and ability to add value to their users. Finally, ambient systems must have anticipatory capabilities to allow them to predict a user's future needs, emotional and physical states. Intelligent urban systems will require sophisticated, self-regulating, quasi autopoietic machine learning algorithms to power the next level of AI that will be required. Kirwan and Fu (2020) introduce "*ambient connectivity*" to describe the ability of AI-driven systems to "sense" their environment in order to continuously optimize network bandwidth and frequencies for optimal energy use and efficiency.

3.4 Anticipatory Capabilities and Biomimetics

The expanded use of technology and data-driven methodologies will allow us to better understand the natural environment and relationships between elements that comprise our intricate planetary ecosystems. In solving some of the most complex challenges, humanity must continue accelerating the pace of innovation achieved through the convergence of technologies. Building resilient systems will require a highly innovative and integrated approach, incorporating a broad spectrum of technologies and capabilities, including sentience, biofeedback, real-time analytics, machine learning, neural networks and artificial intelligence. The more systems converge, the richer the data environment becomes, enabling more accurate predictive models for decision-making. Within the context of a total convergence of systems, these decisions will be made by a new form of collective intelligence. As we continue to expand our knowledge of the natural world, with its intricate design, and elegant adaptations and responses to problem solving, we will more closely align with the natural order. Interpreting natural mechanisms allows us to effectively imitate nature to drive innovation. This approach to design and engineering is known as biomimetics. Within the context of sustainability and autopoiesis, nature's mechanisms are geared towards maximum efficiency with minimal resources. Anticipatory systems contain internal predictive models of themselves and their environment and as such can increase their resilience and chances of survival.

3.5 Self-Regulation, Recursive Interaction, and Feedback

Self-regulation is a key property of autopoiesis that enables systems to consolidate as complex networks. This property can be observed across various classes, scales and domains of systems. It is one of the single most important enablers of life in its purest form. All processes taking place within living systems ultimately support the process of autopoiesis within that system. By removing information asymmetries between different domains, systems and system classes, we improve our ability to integrate, manage and control these entities.

Interactions between systems are seen as mutual disturbances in the sense of each interaction being a stimulus which must be responded to by the interacting systems in a recursive way. Each disturbance will only trigger responses in each respective system that is determined by each system's own organization. Thus, as human beings the changes we undergo as we interact with others in our environment are determined from within by our own unique organization (Leyland 1988, p. 360).

To overcome the limitations of human development originating from the progression of human thought, action and ingenuity related to our complex evolutionary codependency with our environment, new ontological constructs are required to rebalance and redirect the human trajectory and impact on the planet. A collective order based on recursive collective interaction will be necessary to establish this new organizational structure and system of governance. In convergence theory, collective intelligence can represent the fusion of human, technological and natural systems**.** In the form of continuous recursive interaction, real-time metabolic data and biofeedback from a diverse spectrum of human, technological and environmental sources will build a collective awareness and new self-regulating organism. This hybrid organism will naturally require a new operating system, principles and protocols to achieve and sustain autopoiesis.

4 Application: Smart City Operating Systems, Architecture, and Functions

4.1 Smart City Convergence

The concept of smart cities offers a vision for a new paradigm of urban systems that leverage the power of AI-enabled technologies to achieve anticipatory, adaptive and self-regulating abilities. The development and deployment of both human and machine methods and applications incorporating real-time biofeedback will further accelerate the ability for cities to respond to an increasingly volatile world facing numerous threats including climate change, energy crisis, resource depletion, food insecurity, growing populations, infectious diseases and many others. Humans have exacerbated our interrelationship with the natural world, but advancements in neural networks, AI, biomimicry, and generative design, among others, will allow us to develop intelligent self-improving solutions that if harnessed strategically, can bring us once again in harmony with natural systems. This new form of collective intelligence, a combination of the physical city, human organizational behavior and machine intelligence will enable a deeper awareness and symbiosis with other life forms. This total convergence, if applied across all dimensions of urban systems, will allow cities to evolve as autopoietic systems, becoming an integral part of nature itself rather than remaining human-made constructs that deplete nature.

Cities are a part of broader ecosystems including the physical landscape and the natural resources that cities are dependent on. These ecosystems also include the interconnections to other cities and systems that are linked via geographic boundaries, transportation, utilities and telecommunication networks. Through this interconnectivity, cities today are now sharing many of the same technological systems and tech enablers that are propelling cities to become smart. This process of evolutionary convergence is lessening the disparities of cities around the world, simultaneously creating more common autopoietic characteristics that will be needed in the design and implementation of smart city operating systems required to manage the evolution of cities supporting a sustainable planet.

4.2 Autopoietic Operating Systems (AOS)

Cities are complex living organisms in a state of constant evolution, balancing internal and external factors to achieve sustainable growth. To manage the diverse systems and subsystems, there is a need for an integrated city operating system governed by the principles of autopoiesis that we have termed City AOS (Fig. 3). Expanding on the metaphor of the human body, the city operating system can be visualized like the nervous system coordinating the various functions of the city to be harmoniously incorporated within the city. To address the complexity of cities, the design of the City AOS must factor in the specific attributes and

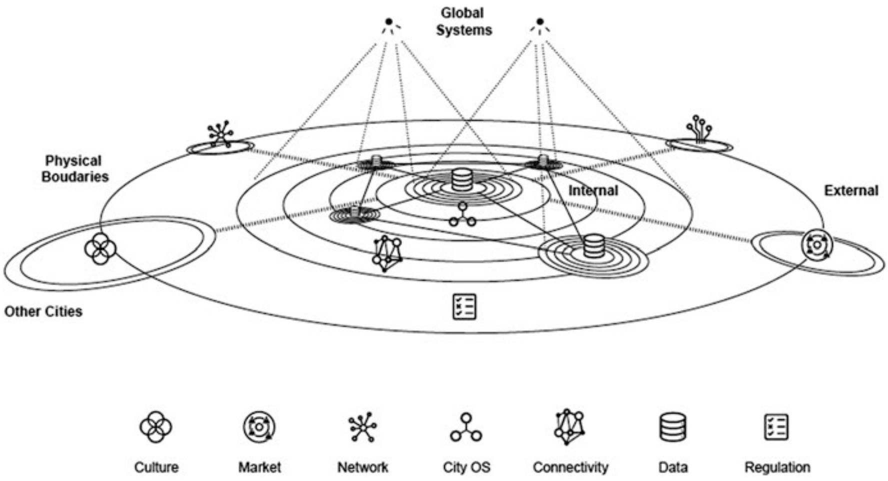


Fig. 3 Autopoietic Operating Systems (AOS) (Adapted from Kirwan and Fu 2020)

requirements of each unique city in order to implement and deploy the appropriate solution. These factors include the scale and configuration of cities from megacities to towns and villages, the governance, cultural, socio-economic composition, as well as the evolutionary trajectory of each city from historical cities with existing legacy systems and infrastructure to emerging cities with the potential to leapfrog with new emerging technological infrastructure.

Previous operating system architecture has typically followed a cybernetic command-and-control or top-down system approach. This has limited operating systems to embody more open, adaptive, living organism behaviors embodied in the autopoietic characteristics we have presented that are in themselves in a state of continual transformation and evolution. Therefore, emerging operating systems will require the exploration of new approaches to the formal language of system architecture representing cities as dynamic living organisms made possible by new human and machine convergence: a hybrid configuration of centralized, decentralized, and distributed systems, deployed across physical and digital dimensions and global, regional, local and hyper-local scales.

4.3 City AOS Architecture

In addressing the complexities and challenges we are facing, the formal nature of operating systems' architecture must be developed based on a new autopoietic language of evolution, adaptation, and transformation. This new architecture should reflect the dynamic flow of the convergence of human, environmental and technological patterns representing the evolution of cities as living, dynamic organisms with the ability to adapt and integrate new technologies within the operational framework and unique culture of each city as a self-generating operating system.

Developing a deeper understanding of city OS requires a clear identification of the system's specific purpose. As a basic rule of thumb, the operating system architecture should be designed according to "form follows function" design principles rather than an imposed formal hierarchy. Like biomimicry, "form follows function" embodies the inherent behavior of the system. The individual and overlapping functions of the city including transportation, energy, security, and communications, each require the development of a suitable operating system architecture that responds to the specific requirements of each function while accommodating the unique DNA of the city as a whole. A key design challenge for system architects, engineers and managers is how city OS should be developed to factor in these diverse functions while simultaneously being designed to evolve and adapt over time, enabling the integration of new improved functionalities and applications that embody and reflect the unique nature and characteristics of each city.

4.4 Meta-Architecture

The meta-architecture establishes the guiding principles and policies that govern the overarching system behavior and determines the structural framework of the system architecture itself. Developing a meta-architecture based on the convergence of the five autopoietic characteristics we defined allows the creation of a system architecture that follows the structural and behavioral principles of living organisms. Studying human anatomy and organic behaviors and patterns found in nature informs the creation of an organizational language and code to describe complex states such as recursive interaction, anticipatory control, adaptation, and self-regulation in the development of new biomimetic system architectures. The ultimate purpose of the meta-architecture is to inform the design and development of an optimal, efficient, flexible and adaptable operating system architecture based on the principles we have defined as autopoietic meta-convergence.

The need for constant upgrades, improvement and optimization of the system necessitates the requirement of a feedback system built into the city OS. In the discourse of meta-design, a related branch of meta-architecture, consideration for how systems will evolve and adapt has been factored into the design process as a given that future uses and problems cannot be completely anticipated during the initial design process. As a response to this criteria, meta-design promotes system architecture design that must have the ability to be adaptive, interactive and self-generative as a living system.

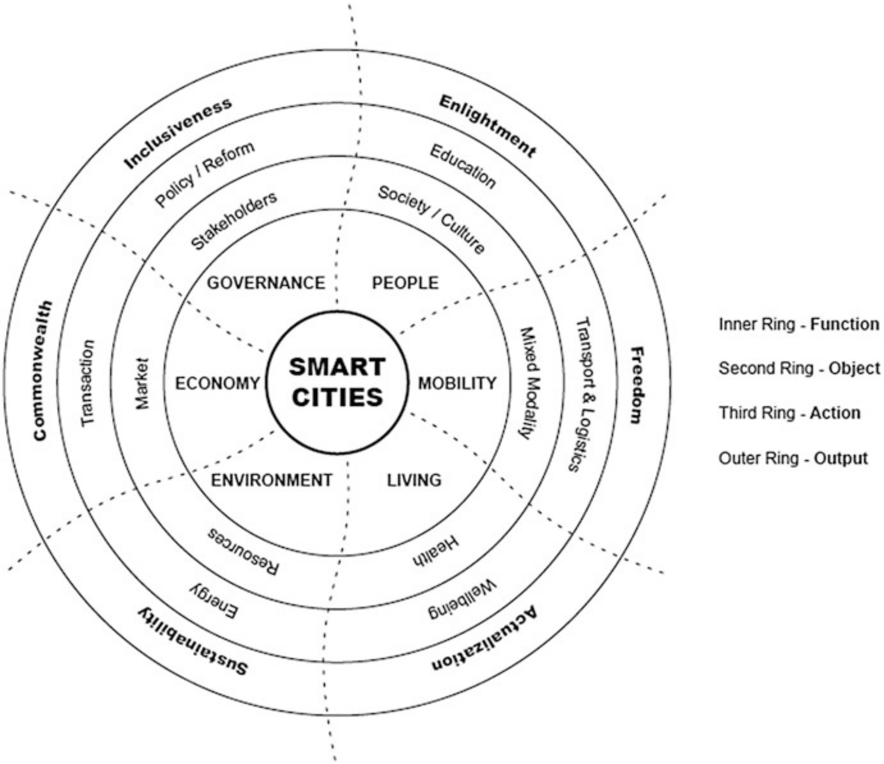


Fig. 4 Smart city functions, enablers, & outcomes (Adapted from Kirwan and Fu 2020)

4.5 Smart City Functions, Enablers, and Outcomes

Cities comprise an extensive, interconnected, and interoperable network of systems and subsystems that support core functionalities and the total life of the city. Each of these functions manifests different autopoietic characteristics and thus have a varied ability to operate within the whole like the human body, which has multiple organs that function independently and collectively as one unified system. To understand and delineate these individual functions within a continuous, connected system—the operating city, the traditional symbolic mandala (Fig. 4) has been co-opted by diverse planners and organizations to assist in explaining these functions in a simple, unified way.

"Borrowing from the visual, spatial and organizational language of the mandala, the city framework can be conceived as a complex set of elements combined within a harmonious whole. Originally, a mandala is the ritualistic creation of an artistic 'circle' intended to represent the universe, a symbolic tribute to the complexity of systems within systems. In principle, a mandala is universe and form combined,

uniting microcosm and macrocosm, the user's experience, and the smart city itself." (Kirwan and Fu 2020, p. 80).

The following descriptions identify the key smart functions of the city and their corresponding autopoietic characteristics as defined in the previous section with the ultimate outcome to be achieved through the meta-convergence of these characteristics.

4.5.1 Smart Environment (Sustainability)

The essential purpose of the Smart Environment is to achieve sustainability representing the balance between the city (the organism) and the environment (the medium). Embodied in the autopoietic characteristics of structural coupling and codependency as natural states of living organisms, neither the organism nor its environment depletes the other and mutual symbiosis is what sustains life. In the case of cities, smart cities must be developed to utilize natural resources in a circular means to build in renewable solutions allowing cities to self-generate rather than parasitically exploit or damage the natural environment to which cities are inextricably linked.

As exhibited in fractals, apparent chaotic and random systems in nature reveal complex, repetitive and predictable patterns. Developing the ability to understand such patterns will allow the virtualization and accurate modeling of the most complex systems in the form of urban ecosystem biomimetic simulation— a multi-dimensional dashboard on which new methods, models, operational logic, technologies and solutions can be tested as living experiments. Biomimicry is about function following natural evolutionary form. The deployment of ubiquitous technologies across a diverse range of regional, local, and hyper-local scales and environments will enable cities to better manage urban metabolism. Over time, the accumulation, aggregation, and convergence of human, machine and environmental data via embedded technologies, smart connected objects and ambient intelligence networks will support new predictive intelligence and regenerative response capabilities to better align with environmental fluctuations enabling cities to harmoniously coexist with nature (Table 2).

Table 2 Smart environment: outcome, autopoietic characteristics, convergent properties, and AOS enablers

Outcom	Autopoietic characteristics	Convergent properties	AOS enablers
Sustainability	Sentience and cognition Ambient intelligence Structural coupling	Anticipatory Biomimetic Ambient connectivity Real-time adaptation	Ubiquitous computing Simulation and monitoring Pattern recognition Satellite/GPS

Table 3 Smart economy: outcome, autopoietic characteristics, convergent properties, and AOS enablers

Outcome	Autopoietic characteristics	Convergent properties	AOS enablers
Commonwealth	Anticipatory capabilities Structural determinism	Transparency Flow Decentralization Resource optimization Peer to peer	AI & ML Blockchain/Holochain Cryptocurrencies Robotics Distributed apps (DAPPS)

4.5.2 Smart Economy (Commonwealth)

The goal of a Smart Economy is to establish a world order that realizes a state of commonwealth for the equitable distribution of resources to all of humanity with a broader planetary benefit. A new convergence of monetary systems and forms of exchange based on decentralized and distributed digital monetary systems will allow the flow of capital in more fluid and targeted means. The ability to localize and scale economic instruments will more effectively serve the needs of diverse and specific populations and uses, protecting natural resources and reducing poverty worldwide. This form of structural deterministic exchange will adapt to meet the demands of each market environment, establishing new forms of economic valuation whereby the true social and environmental cost will be accounted for and all stakeholders including living creatures and nature itself will be recognized and protected for contributing to the maintenance of the ecosystem.

Digital transformation is a key driver for the establishment of new economic systems. Distributed and decentralized platforms, public private partnerships, B2B, B2C, and peer-to-peer (P2P) channels are enabling new forms of value creation and exchange to fully democratize access to resources, markets, and assets. Innovation built on the maximization of data and the continuous convergence of decentralized networks are forming the foundation of a new and exciting economic order defined by principles of sustainability, inclusiveness, social responsibility, societal well-being and commonwealth. This data rich economic order will be defined by an exponential growth in productivity resulting from increased levels of automation and artificial intelligence. Blockchain and Holochain will record transactional patterns as distributed ledger systems securely, efficiently and transparently. Cryptocurrencies and tokenization will offer alternative monetary instruments for exchange and global stores of value (Table 3).

4.5.3 Smart Mobility (Freedom)

Smart Mobility is about a new freedom of choice. The shift from cybernetic-driven point-to-point transportation to a new stage of integrated mobility solutions driven by AI is characterized by multimodality, autonomy and on-demand customized

Table 4 Smart mobility—outcome, autopoietic characteristics, convergent properties, and AOS enablers

Outcome	Autopoietic characteristics	Convergent properties	AOS enablers
Freedom	Sentience & cognition Ambient intelligence Anticipatory capabilities	Optimal pathways Real-time processing Sensorial Autonomous Multimodal	Neural networks AI & ML Smart connected objects Cloud computing Lidar GPS navigation Sensors User-experience Interface

passenger experiences. The transformation and evolution in urban mobility bring an increased sense of well-being for individuals and society. Yet the essence of mobility has not changed much, as people have an intrinsic necessity to be free to explore new destinations and to overcome constraints of distance and time. The self-regulating nature of smart mobility systems is further enhanced by the shifting focus towards creating unique and carefully designed user experiences and lifestyle options as today's citizens become more environmentally sensitive and health conscious and as new forms of travel, spatial-motion, entertainment, and lifestyle converge.

New mobility solutions have been made possible through the convergence of hardware and software technologies and applications working together to bring human-like qualities to machines. Sentience and cognition function as a higher intelligence for inanimate objects to have the ability to read the environmental terrain, process complex multi-dimensional datasets, analyze big data to predict urban conditions and patterns to determine optimum pathways and to enable the continuous flow and combinations of mobility alternatives. By gaining a better comprehension of the architecture and behavior at every scale of our complex dynamic urban systems, we can develop anticipatory capabilities that continually improve through machine learning algorithms as a built-in properties for autonomous and self-regulating systems (Table 4).

4.5.4 Smart Governance (Inclusiveness)

The purpose of Smart Government is to arrive at a global state of inclusiveness characterized by citizen responsibility & participation, transparency, shared resources, and open-source communication. These protocols are geared towards promoting individual and collective well-being, security, and prosperity. The digitalization of public services will enable, promote, encourage, and demand increased new forms of citizen engagement. Certain states around the world have pioneered new inclusive systems of governance empowering their citizens to build on the social organizational theory of self-governance. Rethinking government systems as a series of interconnected cybernetic, autopoietic, anticipatory systems will

Table 5 Smart governance—outcome, autopoietic characteristics, convergent properties, and AOS enablers

Outcome	Autopoietic characteristics	Convergent properties	AOS enablers
Inclusiveness	Ambient intelligence Anticipatory systems	Participatory Collaborative Decentralized Top-down/bottom-up Semantic Self-governing	G2C Interface Civic engagement platform Blockchain voting Real-time streaming APIs Big data analytics AI-augmented government

help governments to better navigate the increasingly complex and decentralized landscape.

The IoT is enabling unprecedented insights into granular data collected from trillions of interconnected devices across cities. This is powering machine learning and consequently enabling increasingly complex predictive modeling used for timely and impactful decision-making. Smart Governance systems, due to their inherent complexities and legacy "code," will require a truly hybrid approach, where all systems must lead to, and over time converge. Predictive analytics and neural networks will empower governance systems to anticipate risks, ensuring continuous improvement and stewardship for citizens. In this way Smart Governance can reach self-regulation through a combination of top-down and bottom-up states in which the awareness and prediction of patterns and trends will optimize system performance. By tapping into collective behavior, the integration of disconnected agencies into a unified platform will enable hyper personalized citizen experiences, bringing about accountability, transparency, and increased trust between governments and citizens (Table 5).

4.5.5 Smart People (Enlightenment)

In the convergent vision of future smart cities, the aim of Smart People will be defined by the pursuit and obtainment of a state of enlightenment for both individuals and the society at large. We define enlightenment as the elevated perception and understanding of ourselves, our social systems and values in the context of developing an advanced awareness and appreciation of the natural environment supporting our existence, sustenance, and prosperity. In the process of realizing this ideal state, the relationship of humans, technology, and nature as an evolutionary structural deterministic bond will have a major influence and impact on the ability of humans to achieve the highest potential both individually, and collectively as a global society. Through technological augmentation, the human ability to tap into the universal spectrum of knowledge, deep cosmic wisdom,

Table 6 Smart people—outcome, autopoietic characteristics, convergent properties, and AOS enablers

Outcome	Autopoietic characteristics	Convergent properties	AOS enablers
Enlightenment	Sentience and cognition Ambient intelligence	Shared and collaborative Open-source Self-perpetuating	Neural networks IoT / cloud Big data AR & VR UX/Interface

and purpose of life will empower individuals and communities to reach their full potential while establishing an interconnected global identity.

As a global interface, the internet has provided a domain for a twenty-first century renaissance in science, innovation, education, and more. This massive transformation has allowed, for the first time, people all around the world to be connected in real-time while simultaneously providing shared access to unlimited information. AI-powered mobile apps are further augmenting user experience, enabling people to not only develop new skills and capabilities, but also to close knowledge gaps faster than ever before, sometimes even instantly. People can now speak and understand multiple foreign languages, while augmented and virtual reality technologies allow us to explore cities without ever stepping a foot in them. Furthermore AR, VR, voice, and image recognition technologies will enable people to overcome physical barriers. The evolution and elevation of a shared global societal platform are a critical enabler of a convergent planetary ecosystem that supports individual and collective advancement within a self-governing and self-perpetuating operating system (Table 6).

4.5.6 Smart Living (Actualization)

Building on the notion of individual and social enlightenment, Smart Living refers to a process of evolutionary convergence that enables humanity to reach a higher dimension or consciousness defined as actualization, or self-actualization, a fusion of mind, body, and spirit representing the optimal state of being. Actualization is a form of transcendence, like Zen, where *being* is realized without consciously attempting to arrive. Actualization is oneness with the self, tapping into the flow and rhythms of life, nature, and society as a unified living ecosystem. This optimal state enhances creativity, health & wellness and affirms overall harmony with the self, society, and nature.

Smart Living will be realized through the next stage of technological advancements in AI: neural networks, augmented reality and enhanced user experiences establishing a collective intelligence that merges humans, machines, and nature. This evolutionary convergence of autopoietic characteristics will enable smart living to be achieved through the direct structural coupling of humans with cities and

Table 7 Smart living—outcome, autopoietic characteristics, convergent properties, and AOS enablers

Outcome	Autopoietic characteristics	Convergent properties	AOS enablers
Actualization	Sentience and cognition Ambient intelligence structural coupling	Collective intelligence Conscientious Urban metabolism Integration	Augmented experiences User interface Smart connected objects Deep AI Pattern recognition

their environmental contexts achieved through the development of an autopoietic operating system interface incorporating a combination of more advanced sentient and cognitive abilities and ambient intelligence to realize a mutual symbiosis that unlocks true human well-being, harmony, and balance with natural systems (Table 7).

5 Conclusion: Towards an Autopoietic City

In the first section of this paper, *Theory*, we explore the evolution of intelligence systems theories to define autopoiesis and convergence as a generalized conceptual foundation. In the second section, *Characteristics*, we identify specific attributes and behaviors derived from diverse theories of systems thinking and autopoiesis to formulate higher-level principles that can be applied to the operations of smart cities. The five characteristics we have identified are s*entience & cognition*, *ambient intelligence*, *anticipatory capabilities*, *structural coupling*, and *structural determinism*. We propose that when combined in the form of a meta-convergence, these characteristics will assist in accelerating the development of self-reproducing and self-regulating smart cities. In the final section *Application*, we explore how these characteristics form the basis of a meta-architecture establishing the guiding principles of smart city development. We describe how these characteristics create a coherent autopoietic system logic at every scale from the overall smart city, the smart city OS, and corresponding system architecture to the sub-functions of the city (*Smart Environment, Smart Economy, Smart Mobility, Smart Governance, Smart People, and Smart Living*).

Cities are complex organisms with multiple individual subsystems operating simultaneously. Considering how many diverse elements interact in real-time, it is remarkable that cities can function as they do without more chaos, especially as many cities are continually expanding. Fortunately, in nature, this chaos is managed through positive and negative feedback in which homeostasis is achieved by the ability of a system to allow the constant flow and exchange of energy between the organism and its medium. This dynamic is part of evolution itself, which is why closed systems are prone to entropy. Applied to the development of future smart cities, the symbiosis of the city and its environment as a form of structural coupling, regulated through biofeedback mechanics that mediate the urban metabolism, sustains a balance of diverse complex functions.

As smart cities evolve, future operating systems will need to develop attributes beyond human intelligence, which has limited our ability to sense the natural environment and to achieve a state of holistic living in harmony with nature. This duality of humankind and the natural world requires the convergence of human, machine, and nature's intelligence to form a higher collective intelligence as a built-in sentience and cognition that manages the complexity of cities, planetary change, and the evolution of technology itself. With the establishment of anticipatory capabilities processing massive real-time human, environmental, and machine data, augmented via AI, an entirely new intelligent, interactive, responsive operating system can emerge as the result of the total meta-convergence of the autopoietic characteristics we have identified.

To address the dynamic nature of operating system, meta-design provides an approach to systems design that incorporates flexibility and adaptability as core design criteria. Smart city operating systems incorporating the underlying properties found in living organisms will allow self-producing, self-healing capabilities. These biomimetic features can be deployed at the city level and across each function, to direct each urban system towards its highest-level state, defined as optimal *outcomes* in the Smart City Mandala model: Smart Environment - *Sustainability*, Smart Economy - *Commonwealth*, Smart Mobility - *Freedom*, Smart Governance - *Inclusion*, Smart People - *Enlightenment*, and Smart Living - *Actualization*.

To arrive at a new convergent autopoietic state, we must overcome the limitations of our superiority complex that we as humans can direct nature to meet our needs, which until now has led in many ways to the severe decoupling of cities from their environments. This fragmentation has been further compounded by the development of smart cities based on top-down, cybernetic command-and-control systems that have mechanized and compartmentalized urban functions. Autopoiesis opens the door for living systems to an entirely new way of approaching the organizational behavior of abstract and organic systems. In the new paradigm described in this paper, understanding the theory of autopoiesis, and applying its converging characteristics is key to unlocking the next level of human innovation in systems thinking for the design and development of autopoietic smart city operating systems (AOS) of the future.

References

Dainow, Brandt. (2017). Smart City Transcendent -Understanding the smart city by transcending ontology. Orbit. 1. 10.29297/orbit.v1i1.27.

Damiano, L. & Luisi, P. (2010). *Towards an Autopoietic Redefinition of Life*. Origins of life and evolution of the biosphere: the journal of the International Society for the Study of the Origin of Life. 40. 145-9. https://doi.org/10.1007/s11084-010-9193-2.

Dunsire, A. (1996). *Tipping the Balance: Autopoiesis and Governance.* Administration and Society, Vol. 28, No 3, November 1996, 299–334.

Geyer, F. (1995). *The Challenges of Sociocybernetics*. Kybernetes Vol. 24, No. 4., pp. 6-32, Netherlands Universities Institute for Coordination of Research in Social Sciences, Amsterdam.

Geyer, F. (2001). *Sociocybernetics*. International Encyclopedia of the Social & Behavioral Sciences, Pergamon, pg. 14549-14554).

Hilpold, P. (2011). *WTO Law and HumanRights: Bringing Together Two Autopoietic Orders*. Chinese Journal of International Law (2011), 323–372.

Kirwan, C. G. & Fu, Z. (2020). *Smart Cities and Artificial Intelligence First Edition*. Elsevier Inc.

Leyland, M.L. (1988). *An Introduction to Some of the Ideas of Humberto Maturana*. Journal of Family Therapy 10: 357-374. The Association of Family Therapy.

Luhmann, N. (2012). *Theory of Society*, Vol 1. Trans. Rhodes Barrett. Stanford: Stanford University Press.

Louie, A.H. (2010). *Robert Rosen's Anticipatory Systems*. Foresight Vol. 12, No.3, pp. 18-29, Q Emerald Group Publishing Limited.

Maturana, H. and Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht/Boston: D. Reidel Publishing Company.

Mele, C., Pels, J., Polese, F. (2010). *A Brief Review of Systems Theories and Their Managerial Applications.* Service Science 2(1-2):126-135.

Mingers, J. (1991). *The cognitive theories of Maturana and Varela*. Systems Practice 4, 321.

Rosen, R. (1978). Fundamentals of measurement and representation on natural systems. General Systems Research Series. New York: Elsevier.

Rosen, R. (1985a). Anticipatory systems. philosophical, mathematical and methodological foundations. New York: Pergamon.

Rosen, R. (1985b). Organisms as causal systems which are not mechanisms: an essay into the nature of complexity. In R. Rosen (Ed.), Theoretical biology and complexity: three essays into the natural philosophy of complex systems (pp. 165–203). Orlando: Academic.

Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, Massachusetts: MIT Press.

Digital "Vitalism" and its "Epistemic" Predecessors: "Smart" Neoteric History and Contemporary Approaches

Konstantinos Moraitis

1 Introductory Remarks: Present "Smart" Paradigms and their Past Cultural Equivalents

The use of the term "smart," describing contemporary scientific, technological, and cultural qualities in comparison to a less "smart" or not "smart" precedent scientific and cultural condition, unveils a rather arrogant statement of poor historical judgment. On the contrary, a more sophisticated and historically sensible observation could insist on the fact that important cultural periods of the past also presented their own important "intelligent" background, in their own characteristic, distinctive cultural way. Moreover, it would be wise, for our present evaluation, to recognize that contemporary "smart" success may have significant historic predecessors going back to the past, at least as far as to the nineteenth century.

This is the aspect that our analysis has decided to pronounce, insisting on the argument that important characteristics of our present epistemological and cultural state may find their first initial reference to the historically older formation of scientific concepts and social or political ethics; as those concerning the importance of the time factor in its correlation to many different fields of cultural production and social expression. It is in this final prospect that the concept of "landscape" attains nowadays, as well as during the nineteenth century, a central cultural value, not only in association to nature and environmental sensitivity but furthermore as an all englobing metaphor, connoting conditions of continuous change, conditions of an "animate" state and morphogenesis.

K. Moraitis (✉)
Professor, School of Architecture NTUA, Athens, Greece
e-mail: mor@arsisarc.gr

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_2

What we should have to stress in addition, as a central methodological attitude of our analysis, is the use of the adjective “cultural,” frequently repeated in the previous lines, as a supplement to the terms “epistemological” or “scientific.” Thus, we would like to indicate that leading scientific paradigms are usually associated with the general cultural production of their societies, imposing their central gnoseological status to many other domains of social expression, to various scientific fields or to the technological and artistic creativity, in a rather unconscious mimetic way. Or, vice versa, generalized beliefs or expressive tendencies may offer the first sparkle, again in an irrational mimetic way, inspiring sciences and arts. If such a description of the development of knowledge and social creativity is not valid, then we have to admit that our presupposed rational scientific production has always been correct and finite, not clouded by the lack of a complete understanding of reality, by ideological delusions and fallacies. Then cultural intelligence could be considered as an unalterable state of affairs, once and for all attained; our “smart” present condition could thus be undeniable for the rest of the historical future. Dialectics of knowledge seem, on the contrary, to present a turbulent, unstable condition of development. Scientific certitudes of the present time may collapse in the future and awareness of reality appears to correspond, in many ways, to the changes of the social, cultural, and political formations.

2 Discussing Neoteric History: Epistemology of Species Evolution, “Smart” Political Theory, and Animate Landscape

Romanticism, at the end of the eighteenth century and later during the nineteenth century was intimately correlated to the veneration of nature. Biology became the leading scientific direction with extended influences to other sciences, natural and social as well. Thus biology offered an extremely all-over cultural paradigm, for arts of representation and construction.

Art Nouveau was a holistic design approach concerning nature-oriented references, as proved by architectural projects such as the interior of *Hôtel Tassel* by Victor Horta (1894) and the *Grand Palais des Beaux-Arts* in Paris (1897–1900), or the object design of the Art Nouveau furniture, or Alphonse Mucha’s posters. In 1847 Joseph Paxton offered to British society the *Birkenhead Park* and, later on, he designed a huge greenhouse, where a great international exhibition, promoting the political, economic, and cultural vigor of Victorian England was sheltered. It was the *Crystal Palace*. Its structural formation was designed, by Paxton, as an imitation of the organic structure of the giant Amazonian water-lily, “Victoria Reggia.”

Historians inform us that, during the summer of 1816, the famous British poet Percy Bysshe Shelley and his lover Mary Wollstonecraft Godwin, later Mary Shelley, visited the Alps, expressing their interest in the mountainous landscape. In the interior of the Mont Blanc Mountain, Percy Shelley described, its “blood”

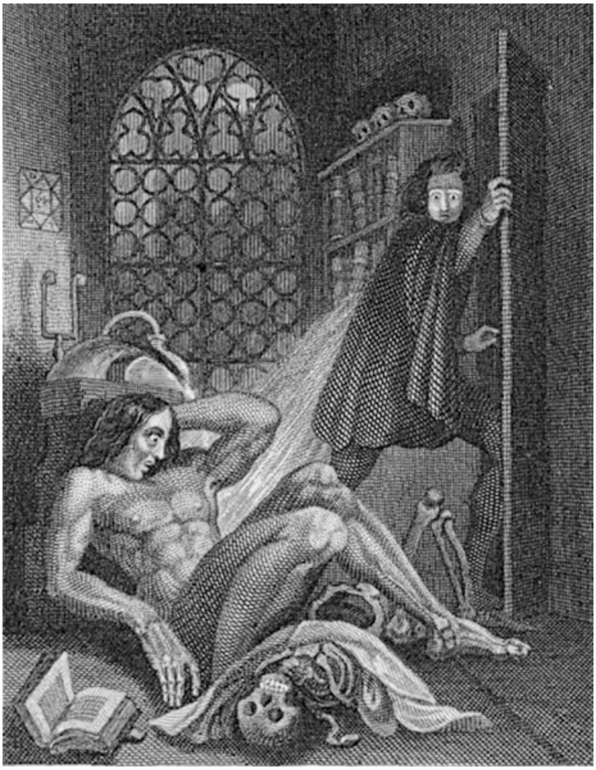


Fig. 1 First comment; nineteenth-century bioethics of body animation. Illustration by Theodor von Holst, from the frontispiece of the 1831 edition of Mary Shelley's *Frankenstein or the Modern Prometheus* (Image in Public Domain https://en.wikipedia.org/wiki/Frankenstein#/media/File:Frontispiece_to_Frankenstein_1831.jpg)

was circulating ... It was during the same period that the German romantic painter Caspar David Friedrich depicted scenes of nature full of inner forces, as *The Wanderer above the Sea of Fog* (1818) and *The Sea of Ice*, also called *The Wreck of Hope* (1823–24). In an analogous way, Mont Blanc was presented by Shelley as a living entity, "a living being" whose "frozen blood forever circulated slowly thro' his stony veins."[1] Some days later his lover Mary would describe a "sparkle which could infuse existence" in doctor Frankenstein's soulless monstrous creation. We refer to her famous novel *Frankenstein or the Modern Prometheus* (Fig. 1) (Shelley 1992, 1994). We may associate the previous reference, concerning Mary Shelley's Frankenstein, with the term "bioethics," with the demand of ethical principles controlling our biological or medical intervention in living organisms; or we may consider that, moreover, they express the belief that organic and inorganic beings, both of them, may acquire common "vital" qualities, able to give life to the organic body as well as to the landscape under morphogenetic process; able to give "form" to buildings in accordance to their "function," in analogy to Charles Darwin's theory of the morphogenetic processes that trans-form animal species.

[1] As presented in Appendix C, in Mary Shelley's *Frankenstein* [1992, p.232].

"Many long discussions took place between Lord Byron and Shelley ..." Mary commended in her introduction of Modern Prometheus. "During those discussions, they referred to the nature of the principle of life and to the possibility to discover it and transfer it to lifeless bodies. They discussed experiments analogous to those of Erasmus Darwin's ... Probably a cadaver could revive ... Galvanism had given indications for the possibility of analogous processes: maybe the component members of a body could be constructed, assembled, and absorb animal warmth."[2] We may explain that Erasmus Robert Darwin (1731–1802), grandfather of Charles's, was a key thinker of the Midlands Enlightenment, well known for his book *Zoomania*. We could also mention Jean-Baptiste Pierre Antoine de Monet, chevalier de Lamarck (1744–1829). According to Lamarck's theory, living species anatomy is created in accordance with functions; thus form is transformed through the use of organs. Later on, Charles Robert Darwin (1809–1882) in his book *The Origin of Species* would suggest that minor differentiations of animal form offer to a part of the population of an animal species the possibility to function in a "better" way, thus to be able to be reproduced in a better way.

The previous Darwinian hypothesis is rather well known. What is less acknowledged is its correlation and impact on social sciences and humanities as, for example, to political philosophy. In 1873 Karl Marx sent to Charles Darwin a copy of the second German edition of the first volume of *The Capital*, with the hand-note "on the part of your sincere admirer," probably indicating a correlation of his theory of dialectic materialism to that of Darwin's. It was in the same context of reference that Marx's close political companion Friedrich Engels described, during the funeral oration dedicated to his colleague, the theory analogies of Marxism to the Darwinian concept of species evolution (Droukopoulos 2018). What was just affirmed, is the overall cultural interest of nineteenth-century for the importance of time and change, as a constitutive condition of natural and social reality; and moreover the importance of biology as a leading paradigm of the overall "epistemic" condition of the neoteric Western world.

3 A Contribution to the Psychoanalysis of the Objective Knowledge: Cultural Inventiveness as a Metaphor of Scientific Paradigms

Trying to analyze the term "epistemic" we may explain that according to Michel Foucault, in his *Les Mots et les Choses—The Order of Things* (Foucault 1966, 1970), it describes the overall cultural field where, at every historic period, knowledge emerges from. Moreover, we could claim, using the previous historic references of the nineteenth century as a characteristic example, that during

[2] As described in Mary Shelley's introduction in the1831's publication of her *Frankenstein* (1994, p. 13).

important periods of constitutive cultural changes, some principle tendencies may be detected which seem to be responsible for common characteristics, in many different cultural domains of the era and thus for its overall identity. They seem to represent not only the central scientific paradigms but also a much larger cultural extension of beliefs and ideological tendencies, the latter not necessarily accepted in a rational, conscious way. They represent as well, different modes of social expression and even with political demands that may merge with what we normally approve as scientific certitude. It was under an analogous orientation that Gaston Bachelard referred to his epistemological analysis using the terms "*contribution à une psychanalyse de la connaissance objective* – contribution to a psychoanalysis of the objective knowledge" (Bachelard 1938), while Althusser (1968, pp. 25–26) tried to merge epistemological approach with the political condition of a given historic period . In our sort and dense reference to the cultural mutations of the late eighteenth and nineteenth-century, previously presented, it is important to mention that scientific assumptions of the natural sciences were merged together with poetic fantasies, literary fiction, and political proposals.

We may ask the help of psychoanalytic theory and explain that according to Jacques Lacan's proposal of the Borromean Knot metaphor (Lacan 1975, p.112), the reality is always presented in osmosis with imaginary views and with the symbolic, semantic formations that offer the structural foundation of the cultural production. In our case of late eighteenth and the nineteenth century, the symbolic order was contributed by the leading paradigm of biology, in accordance to which other theoretical proposals, e.g., political theory, or the creative imagination of other cultural proposals were formed. At the end of the nineteenth century, American architect Sullivan (1896), in his essay "The Tall Office Building Artistically Considered," coined the emblematic for modern architecture phrase "form follows function." "Whether it be the sweeping eagle in his flight," Sullivan described, "or the open apple-blossom, the toiling work-horse, the blithe swan, the branching oak, the winding stream at its base, the drifting clouds, over all the coursing sun, form ever follows function, and this is the law. Where function does not change, form does not change. The granite rocks, the ever-brooding hills, remain for ages; the lightning lives, comes into shape, and dies, in a twinkling." The reference to nature and landscape processes is obvious; the architectural conception of buildings have to imitate what natural examples may teach us. The next paragraph is even more important for our presentation: "It is the pervading law of all things *organic and inorganic*, of all things *physical and metaphysical*, of all things human and all things superhuman, of all true manifestations of the head, of the heart, of the soul, that the life is recognizable in its expression, that *form ever follows function*. This is the law" [1986]. The law according to Sullivan's description is "physical and metaphysical," it possesses the validity of a transcendental principle. We cannot but accept it, as a "naturally" imposed imperative. Moreover, it is a law governing "organic and inorganic" beings as well. This last Sullivan's remark may be compared to Shelley's metaphor about Mont Blanc, according to which the mountain possessed inner life energy; we have cited it previously, accompanied by the term "vital." "Vital" and, in etymological correlation to it, another term "vitalism" that describes a philosophical concept familiar to the romantic German "*Naturphilosophie* – Philosophy of Nature" of the eighteenth and nineteenth centuries. For "vitalistic monism" there is no rigid differentiation between the different states of existence; we may rather refer to a continuous advance from one state to the other. We may thus "perceive mineral as repressed life or rather as potential life; and life in itself, the life of God and spirits included, has to be incarnated or it wishes to: everything in the universe has to attain a form" (Faivre 1982, p. 33). We shall not compare, in this part of our exposition, potential life and movement as presented through vitalism, to the recent "smart" description of animate form which we may reach through topological intuition and computing simulation. Nor shall we insist on the etymology derivation of the word "animation" from the "anima" soul. We shall rather continue by insisting on the religious and theosophical qualities of romanticism and to the comment that the origins of the *Naturphilosophie* and the concept of vitalism are largely associated with medieval and protestant mysticism, with occult intellectuals as Meister Erkhart, Paracelsus or Jacob Boehme. Vitalism may be correlated as well, with the late eighteenth-century spiritualism of Franz Anton Mesmer, insisting on the curative power of "animal magnetism," a natural energy transference occurring between all animated and inanimate objects (Faivre 1982, pp. 41–43). Could we associate Mary Shelley's fictional description of doctor Frankenstein's monstrous creation that came to life during a stormy night, with Mesmer's proposal and the scientific description of electric energy? Could we associate both references with the realization of the neoteric Western society of the eighteenth and nineteenth centuries that its prestigious intelligence largely referred to the scientific successes of the natural sciences and medicine? Certainly yes; however, we have also to admit that those rational acquisitions of the scientific knowledge were also connected with a much broader cultural condition of imaginary projections, in many ways attached to irrational, pre-scientific beliefs, or fictional and artistic metaphors. Moreover, the seductive appeal of those mystic influences did not fade under the predominance of the neoteric rationalism. It is in the introductory pages of his book *The Living City*, that Frank Lloyd Wright cited Paracelsus; "all things are vehicles of virtues, everything in nature is a house wherein dwell certain powers and virtues and such as God has infused throughout Nature and which inhabit all things in the same sense as the soul in men . . . " (Wright 1963, p.vi). We also know that the importance of time, the representational and philosophical principles in the De Stijl approaches were largely relied on Mathieu Schoenmaekers' contribution, the latter being not only a mathematician but also a theosophist. In his book *Du spirituel dans l'art, et dans la peinture en particulier - Concerning The Spiritual In Art*, Wassily Kandinsky (Kandinsky 1989, 1977) exposed the notions which he considered essentials for art, as the "spiritual turning point", "the internal necessity," or the "color." It is significant for this cultural period of the neoteric Europe and Western world that a leading figure of the pedagogical practices of the era, Friedrich Froebel, initiator of the "*Kindergarten"* methods, was supporting a geometrical, abstract figurative didactic attitude, promoting the rational organization of three- and two-dimensional forms. However, he was accepting a quasi vitalistic attitude and believed that common regulatory principles govern the totality of the universe, the astonishing formal structures of the mineral crystals, and the potentials of human mentality as well. It was probably this double attitude of geometrical abstractionism and spirituality that was carried on by Johanness Itten, in his *Vorcurs*, his preparatory lessons for the students of Bauhaus; Itten already possessing a previous didactic experience of Froebel’s pedagogic system. Le Corbusier was also correlated with Froebel’s pedagogy as a child, and probably carried on in his mature creativity, together with the rational clarity of forms the ethical vocation of Purism; pure architectural forms in association with the principles of “clear” thinking.

The previous historic examples present the need of the neoteric intellectuals to encompass main gnoseological trends of nineteenth century; natural sciences appeared to be a first analogous cultural trend of influence, the interest for mental and expressive abstraction, a second one. Thus the theory of schematism, “Schematismus” in Kantian terminology, a part of the *Critique of Pure Reason* (1998, pp. 273–274, A140–143/B179–182) that described mental abstraction as a central process for the creation of scientific concepts and organizational forms, seems to be correlated, as an epistemic cultural tendency, to the major part of modern iconography and architecture. In all previous cases, we could speak of cultural inventiveness as a metaphor for scientific paradigms.

3.1 Cultural Identification With the Leading Paradigms of the Scientific Intelligence

We have already mentioned the French psychoanalyst Jacques Lacan. According to his Borromean Knot concept, individuals have to organize their imaginary intentions in accordance with symbolic orders, which present structures able to organize and offer a state of cultural significance to their imaginative, still not realized desires. Lacan himself “introjected,” in his own theoretical project such a structural system of significance, the topological theory. He was largely influenced by the paradigm of the topological geometry, as, for example, offered by René Thom and his leading mathematical research, presented in books of high international publicity as his *Stabilité structurelle et morphogenèse* (1972) - *Structural Stability and Morphogenesis* (1975). “Introjected,” a term describing the result of the mental mechanism of “introjection,” the way that individuals, or even social groups may unconsciously “identify” themselves with culturally prominent modes of behavior (Laplanche 1978, pp. 209–210 and 187–190). In our case of interest, identification with leading scientific paradigms may mean our desire to appropriate prestigious proposals or, in a disinterested way, our need to imitate innovative concepts and processes that gained our admiration, though we cannot fully understand and assimilate them. In those cases, admiration or desire for cultural prestige appears to surpass the real possibility of understanding, culturally promoting innovative orientations in a number of expressive approaches, outside their initial scientific nucleus.

In any case, we could return to our first assertions, presented previously in our introductory remarks. Accordingly, the cultural intelligence of given societies, during historic periods that present important distinctive characteristics, tends to appear as a multimodal condition, infiltrated in many different scientific and expressive approaches often in an unconscious mimetic way. It thus forms a distinct historical quality, representing the pronounced "smartness" of the era; avant-garde scientific and technological inventiveness together with a general cultural volition for innovative behavior and expression, in many cases presented as an ethical prerogative, a political and moral obligation of the period.

We recently use the word "smart," in order to describe the contemporary cultural association of our societies with the "smart" computational technology; however, this recently introduced connotation does not mean that previous epistemic paradigms did not possess their own scientific and cultural importance, expressed in their own "smart" historically concrete way. Moreover, this recently introduced reference does not mean that the contemporary "smart" paradigm does not possess a scientific and cultural genealogy rooted in previous epistemic predecessors; we have also insisted on this assertion previously. We thus stated that contemporary "smart" mentality may rather be associated with the previously described epistemic condition of the nineteenth century. Even the contemporary notion of the landscape in motion, of "earth that moves" in accordance with the topological intuition and the animate computational simulation, may seek its predecessors in nineteenth-century vitalism approaches, concerning as we already stated, organic as well inorganic formations. In both cases, we may remark that the stability of objects "is eroded by time" (Cache 1995, p. 96).

4 Topological Mathematics, Computational Simulation, and "Folding" Design Forms as a Landscape Reference

Interest for topology was already present during the nineteenth century, with important contributions as the *Vorstudien zur Topologie*, by Johan Benedict Listing (1847) or Henri Poincaré's, *Analysis Situs* (1895). It was during the middle of the last century that the subject attained an extended cultural influence, which however reached its apogee rather recently, after the extended use of computational technology. It is characteristic that the proposal of Gilles Deleuze, who attempted to associate the genesis of topology in Western history with Leibniz and the Baroque period, was largely connected to his academic association with Bernard Cache, whose knowledge of mathematics and even more his immediate experience of computer design applications offered a source of inspiration for the French philosopher. Moreover, soon after the publication of Deleuze's *Le Pli. Leibniz et le Baroque – The Fold. Leibniz and the Baroque* (Deleuze 1988, 1992), folding forms invaded the realm of design practices of every possible scale; in building, landscape, and urban design, as well as in the interior and object design. It is in this context that we may understand the title of Cache's book, *Earth Moves: The Furnishing of Territories* (1995), or the title of its French edition, *Terre-Meuble* (1995, 1997).

In the French language, the world “meuble” refers not to furnishing in general, not to the built-in, immovable pieces, but more precisely to the movable items, which may change their place position. It is in such a metaphoric way that we may approach the earth bas-relief, as a condition under change, as “earth that moves.” In addition, we may approach its design formation, its “furnishing” through computer design in the same animate way that we may approach object design, the design of the objects-pieces of furnishing. It is in this very context that we may also de-codify Cache’s remark to which we already previously referred; “object is eroded by time.” It is in this very context that our design proposals all of them, at any scale of reference, acquired the tendency to become “smart.” They “got the trend” to be correlated to topological inventiveness and computational design applications as well. We mean that since the end of the last century a large part of innovative design “has to be” oriented to those directions, it has to present them in excess, it is “fashionable” to present this “smart” identity in order to be accepted as an innovative creation. Likewise contemporary design approaches often mimetize the topological folding continuity of a curvilinear surface. In their awarded competition proposal for the Yokohama Port Terminal, FOA architects presented a 3D image of a continuous folded form, as produced by their computer parametric approach. In reality, the built final surface of the terminal was not a continuous, “smooth” surface as depicted in the 3D presentations of the competition proposal. It pretended that it was continuous, non-fragmentary, and articulated, though it was formed on a “striated,” metallic bearing structure.

“Smooth” and “striated” are attributes borrowed from Deleuze and Felix Guattari’s *Mille Plateaux - A Thousand Plateaus* (Deleuze & Guattari 1980, 1987) in order to describe, the first one topological continuity versus the second term that is representative of a Euclidean space conception. In both cases that of the Yokohama Terminal and that of Deleuze and Guattari’s use of the terms, we may point out the expressed desire to bring in contact architectural or theory conception with scientific and technological orientations that seemed to be nodal for the cultural identity of the era. Topology and computer technology may be considered as crucial analogous cultural orientations, representing the “smart” predominance of our epoch. It was in correlation to them that the notion of landscape was brought forward, associated not only to environmental sensibility. Furthermore, it was used as a metaphor of the topological morphogenesis, in order to describe design forms whose presentation was possible through computational simulation. It was in this context that analogous building constructions were indexed as “landscape formations,” as by architect Zaha Hadid, connoting rather their reference to topological conception than to “green” sustainable pretensions. In all those topologically oriented architectural proposals, morphogenetic volition was realized more in a figurative way than literally; in a figurative way reminding us of the desperate effort of European modernism to represent time and movement through immovable depictions and motionless sculptural or architectural structures. Likewise, the contemporary claim for cultural intelligence seems comparable to the mid-war avant-garde proposals concerning "space and time" relations; Siegfried Giedion could thus be considered as the predecessor of our contemporary "animate forms" apologists.[3]

4.1 A Space-Time Metaphor Not Realized in Building Construction

Speaking in terms of historical reference, the cultural agony for a time agitated reality attained an important theoretical status long before computational animate simulation. Kantian philosophical inventiveness described space and time, in his *Critique of Pure Reason*, as the inter-related "conditions of sensibility, under which alone outer intuition is possible for us" (1998, p. 149, A26/B42). Nothing could be conceived outside space and time reference. This initial "smart" contribution to neoteric theory could certainly be associated with the general epistemic identity of the eighteenth and nineteenth century, to the previously cited influence of concepts of change and metabolism introduced through biology, political theory, and for sure through the continuous social radical change. Historians also correlate it to the development of the mechanized world, to steam machines, to moving engines of all sorts, invading the state of production and transportation. In a certain way, the everyday reality was redirected on the ground of new scientific and technological proposals; however, we may again insist on our remark about a non-fully conscious development of a culture that is also extended at the field of science. Thus, the sensation of a new mechanical "motivation" attended an extended expressive status. We have already mentioned the concept of "function" that survives in architectural discourse for more than a century, pretending that architectural objects, though largely immovable in themselves, may activate functional uses. In terms of constructive reality, the only practically movable parts of buildings or urban formations are those that refer to mechanical and electromechanical devices. This lack of real movement was culturally replaced in the immovable "body" of architecture or in the immovable depictions, during the first decades of the twentieth century, by the extended formal metaphor of the desired action. It is not strange that the holy bible of architectural modernity bears the title *Space, Time and Architecture* (1967); a correlation of terms that, according to its author, had attended the status of a new "tradition"; a status of cultural influence extended outside centralized conscious expression as established custom and belief. It was in this sense that De Stijl architects and painters were trying to expand three-dimensional spatial stability by expressing the fourth dimension of time; Russian constructivism presented a potential movement, while Vladimir Tatlin attempted to connote the dialectics of history, the Marxist approach of continuous social change, through the twin helixes of his *Monument to the Third International*.

[3] Giedion (1967).

We have just presented a minority of the relevant modernistic representational examples. Italian Futurism was possessed by the agony of expressing the temporal "continuity in space," and Cubism was trying to express different aspects of the same object as perceived by different points of vision ... a "smart" neoteric volition of movement introduced into the representational practices. It was not fully attended as constructional reality; however, it was intensively expressed as metaphoric reference. Such an intense metaphor surely presented the desire, the imaginary cultural need to incorporate innovative approaches to scientific and technological applications, to animate representation and construction, though the latter were not yet ready to attend them. In present-day terms, the topologically oriented folding forms in architectural or object design are also largely metaphoric. They are probably created through the representational transformations of the "animate" computational simulation but they do not really move as final constructional entities.

4.2 A State of "Non-ordinary Reality": Augmented Conscience and Brazilian Hallucinatory Modernity

During the ninety-sixties a general disturbance of Western ethics appeared on both sides of the Atlantic Ocean. In the European continent political confrontation suddenly obtained a quasi-revolutionary presence; in Prague and in 1968 Paris as well. It is important to mention that this political controversy attained the value of an extended cultural demand, largely correlated with youth in general and university students in particular. In Poland, Yugoslavia, France, and the States, "strawberry statements" intensively declared by intellectuals and academic communities were asking for the reevaluation of political ethics and everyday morality in an all-encompassing cultural way. In correlation to their subversive approaches, new aspects concerning the creation and the didactics of sciences appeared. Knowledge could not be regarded as politically neutral; on the contrary "smart" scientific proposals had to coincide with a "smart" new political state of conscience. Nowadays, half a century after the period mentioned, it seems that the volition for an "augmented" state of conscience appeared to be a characteristic quality of the "swinging sixties"; in the case of the political movements as well as in the case of the everyday morals and ethics. Images and forms of behavior, being stabilized for decades were disorganized. They were transformed by the invasion of the fluctuating, hallucinatory appearance of continuous motion. Movable curvilinear shapes of a "yellow submarine" were navigating in the cultural and epistemic ocean of the world. Even the production of chemical drugs as LSD also known as "acid" were often correlated by artists and intellectuals of the era with a "non-ordinary" state of conscience that could produce an augmented vision of reality,[4] similar to

[4] 'A state of non-ordinary reality', as described by the anthropologist and literature writer Carlos Castaneda in his bestseller *The Teachings of Don Juan:* A *Yaqui Way of Knowledge* (1968). It is

the endoscopic traditional practices of meditation, surpassing the restricted rational intellectual orientation of the Western civilization.

We shall not continue repeating that “smartness” and “intelligence” are not scientifically and technologically isolated but rather a state of conscience, probably concerning the recognition of the totality of the social existence at a given historic period. We shall only comment in addition that the curvilinear forms in continuous cinematographic motion, which appeared in the given period of the sixties, may deserve a topological identity. We shall continue by demanding ourselves whether those forms in “smart” movement, which pre-iconized “smart” computer simulation and parametric design, had a common orientation with forms that appeared in the landscape design of Brazilian modernity and in the garden art of the sixties and seventies, namely in the landscape proposals of Roberto Burle Marx.[5]

5 A Potentially “Smart”’ Conclusion: Epigenetic Desire, Animate Design, and “Digital Vitalism”

Could animate design really produce “smart” animate architecture?

The applied technology of computational fabrication seems to present the closest example to this desired “smart” scientific and cultural condition. However what we describe as “animate design” still remains at the state of a premature indication of what we hope to produce as “smart” architecture in motion. Architectural and design formations could probably hope to attain the living status of organic beings, imitating their change of forms and probably their cellular reproduction. The metaphor of earth bas-relief, we have repeatedly referred to it, seems to present the first step of this cultural expectation for “theopiia,”[6] of this volition for divine-like creation. The landscape could produce the first simulation of morphogenetic change. It could even be correlated to “epigenetics,” to the description of heritable phenotypic change in general. Is the description of “epigenetic landscape” as proposed by Conrad Waddington an overall epistemic metaphor, bringing together topological intuition and biological theory ... or, is it a scientific reference to

significant that, though the book content was initially submitted as a scientific text, as a master thesis of the author, it was closely correlated to the general cultural predisposition of a number of artists and intellectuals of the period, rejecting conventional Western ethics and promoting cultural paradigms of *La Pensée Sauvage - The Savage Mind* – (Lévi-Strauss 1962, 1966); a product of scientific orientation associated to the general cultural atmosphere of the era.

[5] Hallucinatory Brazilian modernism and the aesthetics of the animate landscape: Curvilinear geometry, as applied at the garden design proposal of the *Beach House for Mr. and Mrs. Burton Tremaine* by Roberto Burle Marx (site plan 1948 – see “Beach House for Mr. and Mrs. Burton Tremaine, project, Santa Barbara, California” – available from: https://www.moma.org/collection/works/161, 29 June 2021).

[6] “Theo-piia,” a Greek compound word bringing together the term “theos – god” and the term “piia – creation”: divine creation.

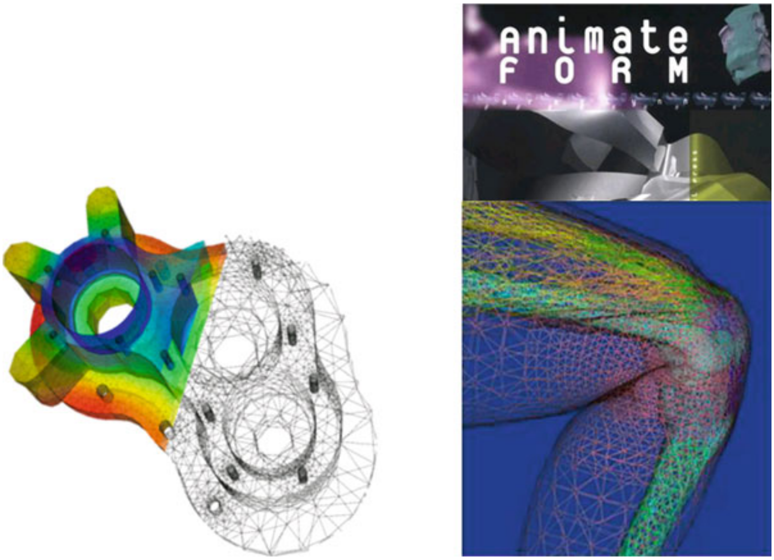


Fig. 2 Third group of comments, concerning contemporary “digital vitalism”: 2.1. image on the left; computational mechanics as applied in machine design (https://en.wikipedia.org/wiki/Differential_equation#/media/File:Elmer-pump-heatequation.png). 2.2. right on the top; Greg Lynn’s book on the computational animate form (reproduction of the frontispiece of the homonym book - Lynn 1999 - http://ka-au.net/animate-form-by-greg-lynn/) and 2.3. right on the bottom; computational simulation as applied in living body analysis (http://www.cs.unc.edu/Research/us/fem/)

effects on cellular and physiological phenotypic traits, which may result from external or environmental factors?

In any case, the utmost contribution of “animate design,” (Fig. 2) as proposed by computational mechanics does not seem to be restricted to objectified structures, to design objects even if they could offer a potential state of movement. Its utmost expectation is to invade in the very core of the living condition, to establish their “smart” inventiveness as “digital vitalism,” able to re-form living structures.

References

Althusser L.: *Pour Marx*. Paris, Éd. François Maspero (1968).

Bachelard G.: *La Formation de l’esprit scientifique. Contribution à une psychanalyse de la connaissance objective*. Paris, Éd. Vrin (1938).

Cache B.: *Earth Moves: The Furnishing of Territories*. Cambridge, MIT Press (1995) – *Terre meuble*. Orléans, Éd. HYX (1997).

Castaneda C.: *The Teachings of Don Juan:* A *Yaqui Way of Knowledge*. Berkeley, Univ. of California Press (1968).

Deleuze G. & Guattari F.: *Mille Plateaux. Capitalisme et Schizophrénie.* Paris, Éd. de Minuit (1980) - *A Thousand Plateaus. Capitalism and Schizophrenia*. Minneapolis, Univ. of Minnesota Press (1987).

Deleuze G.: *Le pli. Leibniz et le baroque*. Paris, Éd. de Minuit (1988) –*The Fold. Leibniz and the Baroque.* Minneapolis, Univ. of Minnesota Press (1992).
Droukopoulos V.: «Μαρξ και Δαρβίνος. Με αφορμή τον μύθο μιας αφιέρωσης – "Marx and Darwin. On the occasion of a legendary dedication". In: *Thesis – Θέσεις,* Greek magazine. Athens, July-September 2018, pp. 83-95.
Faivre A.: «La philosophie de la nature dans le romantisme allemand». In: *Encyclopédie de la Pléiade*. *Histoire de la Philosophie, tome 3: Du XIX Siècle a nos jours*. Paris, Éd. Gallimard (1974) - Greek translation: *Ιστορία της Φιλοσοφίας. 19ος αιώνας, Ρομαντικοί – κοινωνιολόγοι.* Athens, MIET publ. (1982), pp. 25-65.
Foucault M.: *Les Mots et les Choses: Une archéologie des sciences humaines*. Paris, Éd. Gallimard (1966) - *The Order of Things: An Archaeology of the Human Sciences.* New York, Pantheon Books (1970).
Giedion S.: *Space, Time and Architecture: The Growth of a new Tradition*. Cambridge, Harvard Univ. Press (1967).
Kandinsky W.: *Du spirituel dans l'art et dans la peinture en particulier*. Paris, Éd. Gallimard (1989) – English translation: *Concerning The Spiritual In Art*. Toronto, Dover Publ. (1977).
Kant I.: *Critique of Pure Reason*. Cambridge, Cambridge Univ. Press, (1998).
Lacan J.: Le Séminaire, Livre XX. Encore, 1972-73. Paris, Seuil (1975).
Laplanche J. & Pontalis j.-B.: *Vocabulaire de la Psychanalyse*. Paris, Presses Universitaires de France (1978).
Lévi-Strauss C.: *La Pensée Sauvage*. Paris, Librairie Plon (1962) - *The Savage Mind*. London, Weidenfeld and Nicolson (1966).
Lynn G.: *Animate Form*. Princeton, Princeton Univ. Press (1999).
Shelley M.: *Frankenstein or the Modern Prometheus*. Oxford, Oxford Univ. Press (1992) – Greek translation: *Φρανκενστάιν*. Athens, Livani publ. (1994).
Sullivan, L. H.: "The tall office building artistically considered". In: *Lippincott's Magazine*. Philadelphia, J.B. Lippincott, March 1896, pp. 403-409.
Thom R.: *Stabilité structurelle et morphogenèse.* Paris, InterÉditions (1972) – *Structural Stability and Morphogenesis*. Reading, W. A. Benjamin INC, (1975).
Wright F.L.: *The Living City*. New York, Mentor Books (1963).

Unbuildable Cities

Stamatina Th. Rassia

1 Introduction

Ever since the discovery of fire, human beings have been building their societies and structuring their lives based on new technologies. Over the course of the last century, architectural thinking has directed the design of the future living conditions of our societies using new—at each given time—technologies. Many of these designs were deemed unbuildable and others were considered to be so futuristic that in fact were presented as artwork (in MoMa and other famous museums and art galleries). While many of these products of thinking have not been built, interestingly these seem to have formed constructs of future architectural ideologies and implementable designs. Many of the most famous Architects of the 20th century thus seem to have used futuristic or rather more abstract drawings of architectural theorists and designer groups to seek for inspirational points or driving conceptions for their own works (cf. the work of Archigram, Superstudio, and Piano & Rogers at Pompidou). The issue of technological advancements, the discovery of robots and the advancement of machine learning, artificial intelligence and more, have been at the forefront of architectural thinking. However, these have not managed to date, to influence whole metropolitan areas or conurbations. Technology so far seems to empower simple or minimal actions in or around building networks. Hence, the aim of this paper is to start an open dialogue on the future of urban design in its totality, especially now that our contemporary lifestyles face serious public health crises, pandemics and issues of environmental sustainability.

S. Th. Rassia (✉)
University of Macedonia & Le.D.R.A. Group, Greece

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_3

2 Architectural Design and the Machine Ages

One of the initial moments where an architect approached the topic of the machine age in urban environments[1] has been marked by the dialectic and design work of the Italian Architect Antonio Sant'Elia (1888-1916). Sant'Elia followed the early 20th century Futurist movement in Architecture and envisioned that cities should be mechanized and organized as vast, industrialized and interconnected urban conurbations that would support "the life" of the city.

Such ideas started to flourish around 1912 to 1914, as famous architects like Otto Wagner, Adolf Loos and Renzo Picasso (Genoa) followed the example of Antonio Sant'Elia (Fig. 1) and drew surreal and unusual buildings for a so-called Citta Nuova (New City). These drawing manifestations were exhibited in the "Nuove Tendenze" group exhibition in May 1914. This marked a symbolic approach to industrialization in relation with the new age and their works are displayed nowadays in Como's art gallery.[2]

Interestingly, Sant'Elia proposed that conventions and customs should give way to the Modern "gigantic machine" city.[3] In fact he seems to follow the futurist visions of the Italian poet and artist Filippo Tommaso Marinetti by suggesting that classical elegance should be replaced by an "immense, and tumultuous shipyard,"[4] and this should be our new model space for living in.

While population growth led to the physical expansion of urban metropolitan areas, the vast industrialization attracted more innovative thinking. In 1915, the innovative thinker on urban planning and sociology, Patrick Geddes, added to the architectural nomenclature a new term that also constituted the title of his book, called *Cities in Evolution.* The subtitle of this book, where he highlighted the capacity of motorized transport systems to function with new technologies and electric power, was *An Introduction to the Town Planning Movement and to the Study of Civics.*[5] The underlying conception of this movement, as he called it, was to give rise to a higher and more systematic connection between cities that grow.

A responsibly powered architecture based on the use of new technologies that would influence people's lives in a positive way[6] has been also the topic of inspiration for Cedric Price (FRIBA Architect, 1934-2003). Price is described as an architectural optimist, much like the famous American architect and systems theorist, futurist and inventor Buckminster Fuller (1895 – 1983).

---

[1] *Manifesto of Futurist Architecture* ("Manifesto of Futurist Architecture". *Retrieved 11 March2016.*)

[2] McGarrigle, Niall (12 March 2016). "The Futurist world of architect Antonio Sant'Elia". The Irish Times. Retrieved 12 April 2017.

[3] Banham (1960)

[4] Banham (1960)

[5] Geddes (1915)

[6] https://www.independent.co.uk/news/obituaries/cedric-price-36932.html

Fig. 1 Antonio Sant'Elia, Image source:https://web.stanford.edu/~kimth/www-mit/ma s110/paper1/

2.1 Architecture for Flying Cars and Spaceship Earth

One of the most inspiring architects around the globe, who influenced a more humanistic and at the same time experimental and geometric approach to architecture was Buckminster Fuller. From a very young age he expressed his ideas on the ways geometry was taught. He thought that dots on chalkboards represented "empty" points which lines could cross while extending to infinity. He preferred more synergetic approaches to design and he focused on machine or tools' development by using materials like wood. By the age of 12 he developed his own propulsion system for small boats. After his studies at Harvard in 1927, Fuller faced serious survival issues at the start of his career, when in exchange for meals he decorated the interiors of a popular café, called Romany Marie's. On this occasion he met and then collaborated with Isamu Naguchi (Japanese-American artist and landscape architect) and Starling Burgess (American yacht designer and naval architect) for the development of a concept car called the Dymaxion (Fig. 2), that

Fig. 2 Image of Dymanxion car replica (Source: https://en.wikipedia.org/wiki/Dymaxion_car#/media/File:Dynamixion_car_by_Buckminster_Fuller_1933_(side_views).jpg)

was designed to offer a new omni-medium transport[7] able to be driven but also to take-off and fly.

In the effort to encourage a harmonious symbiosis, Buckminster Fuller also coined the term "Spaceship Earth" since he saw the earth as a global unity that required all the people of the world to act together towards a greater good. He demarcated also his interest in sustainability, design replication, material efficiency and harmony with the surrounding environment as well as versatility. In this sense, Fuller developed his thin-shelled geodesic dome, known as the Montreal biosphere, U.S.'s 1967 World Exposition pavilion in Quebec[8]. In developing such dome looking structures, he aimed to make a point about architecture and versatility (see following Figs. 3, 4).

2.2 *Robot Cities*

Between 1961 and 1974, the Archigram group explored how advancements of technology could lead to design scenarios and hypothetical new realities. "Walking Cities, I venture. If you can build an ocean liner, why not them?",[9] Archigram's Peter Cook said in 2018.

The UK based Archigram group, composed by David Greene, Warren Chalk, Peter Cook, Michael Webb, Ron Herron and Dennis Crompton, created a set of drawings where they expressed their views on cities as technologically advanced and rather futuristic machines for living in. They did however seem to dismiss the small-scale and human level to their drawings. These visions, such as the plug-in

[7] Robert (1973)

[8] https://www.archdaily.com/572135/ad-classics-montreal-biosphere-buckminster-fuller

[9] https://www.theguardian.com/artanddesign/2018/nov/18/archigram-60s-architects-vision-urban-living-the-book

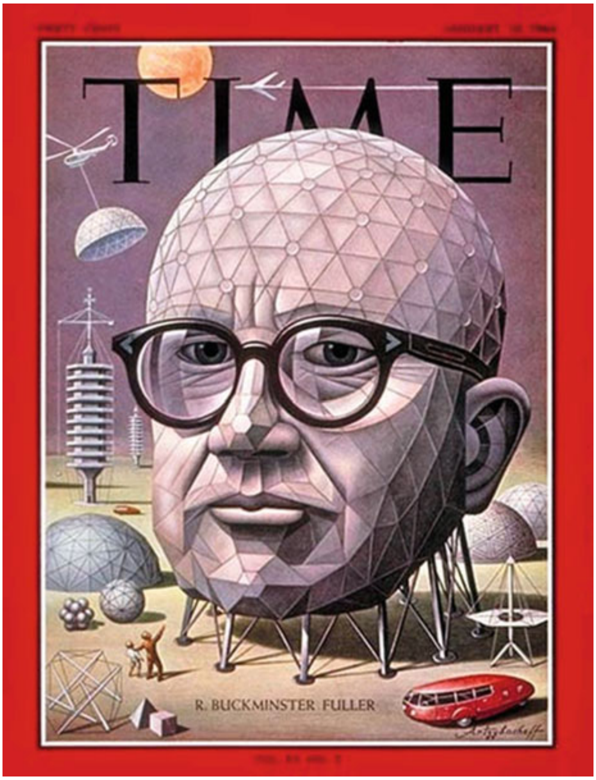


Fig. 3 TIME Magazine January 10, 1964. Image © Time Inc. Source: "https://www.archdaily.com/ 572135/ad-classics-montreal-biosphere-buckminster-fuller"

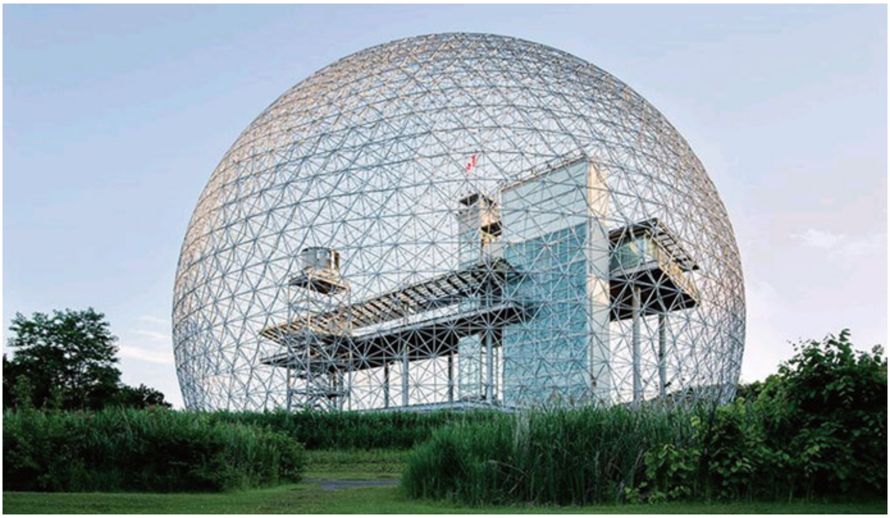


Fig. 4 Montreal biosphere. Image source: https://www.azuremagazine.com/article/buckminster-fuller-montreal-biosphere/

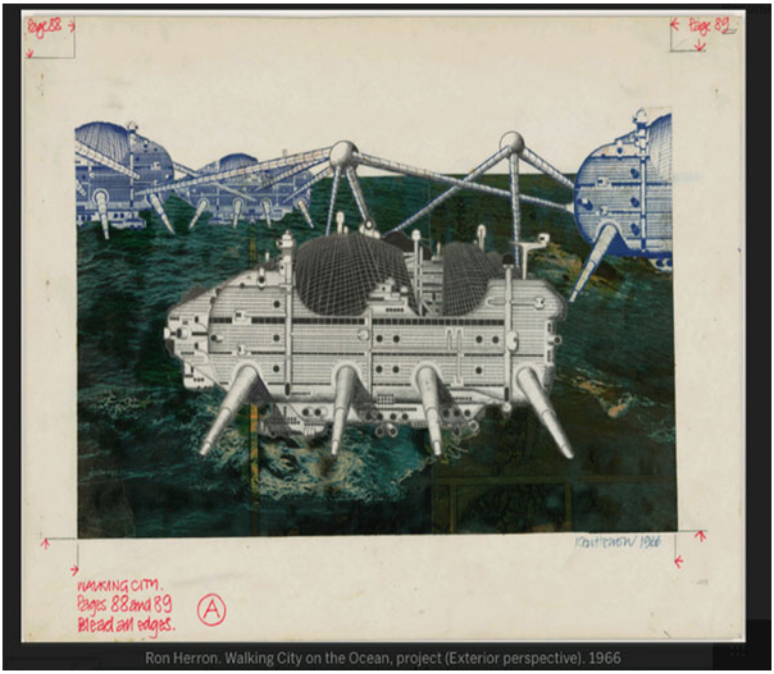


Fig. 5 Archigram Ron Herron's. Walking City on the Ocean (1966). Source: https://www.moma.org/collection/works/814

city, the instant city or the living pod, although deemed as unbuildable, have inspired the work of many famous architects' around the globe, such as Renzo Piano and Richard Rogers as well as Nick Grimshaw, Rem Koolhaas and Will Alsop. More specifically, the Centre Pompidou in Paris was designed as an "evolving spatial diagram",[10,11] and this inspiration was attributed to the architectural inspirations of Archigram and the Superstudio, an architectural firm founded in 1966 in Florence.

Interestingly, Archigram considered that built space had the flaw of being static and thus worked towards suggesting utopian yet fascinating mobile alternatives, where cities were converted into giant insect-looking walking robots that were able to take their resources and goods and plug them into safer environments when needed. Ron Herron's (Archigram member) envisioned the "Walking city" (Fig. 5) as a survival strategy for nuclear wars, where citizens of the world would have to seek refuge in safer places.[12]

[10] https://www.centrepompidou.fr/en/The-Centre-Pompidou/The-Building

[11] https://www.dezeen.com/2019/09/18/archigram-design-trust-gala-2019-hong-kong/

[12] https://medium.com/@emilyrowlings/a-walking-city-archigram-and-ron-herron-7dbf2c8fae99

2.3 Radical Architecture

In parallel to the abovementioned, the Italian team Superstudio (of Adolfo Natalini, Cristiano Toraldo di Francia, Gian piero Frassinelli, Alessandro Magris, Roberto Magris and Alessandro Polli) manifested so-called radical designs[13] in the form of a homonymous movement. They suggested three areas of research in architecture. Adolfo Natalini coined these areas in 1967 as, "technomorphic architecture",[14] "architecture of the movement" and "architecture of the image". Technomorphic architecture belonged to their view of the city as a "Vacation Machine". This architecture viewpoint proposed that we should have no street fronts, as the so-called vacation machine is part of a larger ecosystem, much like Buckminster Fuller's vision of "spaceship earth". In the city as a vacation machine, all buildings would have their "feet in the sea",[15] as written by Cristiano Toraldo di Francia (Superstudio member).

3 Discussion

From Futurism to Archigram's visions and beyond, industrial design in architecture structures lifestyles in machine-like environments. The key question however is whether one aims to see the city of tomorrow as a factory of the future or rather as a means of promoting a new more rational and sustainable lifestyle for its users' health and wellbeing.

Cristiano Toraldo di Francia has suggested that: *"This new ambiguous relationship between nature and architecture alluded to the passage from Hans Hollein's "Alles ist Architektur" to the concept that "everything is landscape": architecture and landscape are no longer distinct entities; architecture can restore the landscape by occupying the non-places -voids resulting from natural actions- recovering interrupted layers for architecture itself. Architecture is reduced to a neutral, transparent surface as a further layer on the surface of Spaceship Earth, just as homes can be dug into the earth or utilize folds, while the whole functions as a diachronic theatre without walls, but with mobile systems. The office becomes landscape, and the landscape enters the museum."*[16]

Interestingly, Superstudio aimed at spurring social change by design. In the words of Adolfo Natalini (1971) *"...if design is merely an inducement to consume, then we must reject design; if architecture is merely the codifying of bourgeois model of ownership and society, then we must reject architecture; if architecture and town*

[13] Didero, M. C. (2017). *SuperDesign: Italian Radical Design 1965-75*. New York: Monacelli Press. p. 210. ISBN 978-1580934954

[14] https://www.cristianotoraldodifrancia.it/the-beginning-technomorphic-architecture/

[15] https://www.cristianotoraldodifrancia.it/the-beginning-technomorphic-architecture/

[16] https://www.cristianotoraldodifrancia.it/the-beginning-technomorphic-architecture/

planning is merely the formalization of present unjust social divisions, then we must reject town planning and its cities...until all design activities are aimed towards meeting primary needs. Until then, design must disappear. We can live without architecture...".[17]

4 Open Question

Over decades, architectural thinking processes have undertaken the idea of advancing technologies and their relation with the city as a unit where societies co-exist. Others faced this as a socio-environmental and cultural question, others as a phenomenological one, and others as a technocratic and a systemic question.

In this period, where vast urbanization prevails and converts lifestyles into metropolitan ones, machine learning and artificial intelligence should be applied wisely on a variety of installations. Cities differ considerably from one another, therefore, holistic design approaches cannot be applied naturally.

While theories and technologies have developed worldwide, the question persists, why our cities have not yet been developed as machines for living-in safely, sustainably and peacefully? While image processing, censoring temperature fluctuations, and data analysis can actuate appropriate model responses to urban calamities, it is worth questioning why mathematical algorithms have not yet empowered an ecosystem of models that could optimize lifestyles over conditions of crisis, such as global pandemics.

References

Banham, R. (1960), "Theory and Design in the First Machine Age". Cambridge, MA: MIT Press.

Didero, M.C. (2017)."*SuperDesign: Italian Radical Design 1965-75"*. New York: Monacelli Press, p. 210. ISBN 978-1580934954

Geddes, P. (1915), "Cities in Evolution: An Introduction to the Town Planning Movement and to the Study of Civics", London: William & Norgate.

Robert, M. (1973), *"The Dymaxion World of Buckminster Fuller ",* Anchor Press: Doublebay, p.104.

Electronic Sources

Manifesto of Futurist Architecture ("Manifesto of Futurist Architecture". *Retrieved 11 March2016.*)

[17] Didero (2017)

https://en.wikipedia.org/wiki/Dymaxion_car#/media/File:Dynamixion_car_by_Buckminster_Fuller_1933_(side_views).jpg
https://medium.com/@emilyrowlings/a-walking-city-archigram-and-ron-herron-7dbf2c8fae99
https://web.stanford.edu/~kimth/www-mit/mas110/paper1/
https://www.archdaily.com/572135/ad-classics-montreal-biosphere-buckminster-fuller
https://www.azuremagazine.com/article/buckminster-fuller-montreal-biosphere/
https://www.centrepompidou.fr/en/The-Centre-Pompidou/The-Building
https://www.cristianotoraldodifrancia.it/the-beginning-technomorphic-architecture/
https://www.dezeen.com/2019/09/18/archigram-design-trust-gala-2019-hong-kong/
https://www.independent.co.uk/news/obituaries/cedric-price-36932.html
https://www.moma.org/collection/works/814
https://www.theguardian.com/artanddesign/2018/nov/18/archigram-60s-architects-vision-urban-living-the-book
McGarrigle, Niall (12 March 2016). “The Futurist world of architect Antonio Sant’Elia”. The Irish Times. Retrieved 12 April 2017.

Smart Cities as Identities

Nikolaos Tsoniotis

1 Us and the City

From the moment we wake up in the morning to the time we go to bed at night, and even when we sleep, there so many little things that take place. We wake up, switch on the bathroom light, have a shower, wash our teeth. We then switch on the coffee maker, have a cup of coffee, wear our clothes, and get out of the building. We get into our vehicle, drive to work, wait at the traffic lights, get into elevators, make phone calls, use a computer maybe, submit an application to a municipal or governmental service . . . and the list goes on and on.

All these actions, whether you realize it or not, at their core need two things to take place either implicitly or explicitly. You and a rule-based agreement. In other words, your personal identity and a set of rules that make sure, whatever you do can be attested somehow. Your national identity (national ID) is typically needed for the contract with the utilities which give you electricity, gas, water, phoneline. It is necessary when you go to the bank to transact, when you interact with governmental services. Aspects of it are needed to enter a place at work with a code, to log into your computer or a web service. Similarly, predefined rules dictate your benefits and obligations, your utilities will continue to serve you if you exchange your money for the equivalent value of the electricity you consumed, the value of which is set based on market rules, time of day, contractual agreement. Your application will be processed if you can prove your identity, your car will be filled up at a gas station if you pay for it and depending on your bank card or loyalty agreement you will probably get some loyalty points.

N. Tsoniotis (✉)
Ideas Forward P.C., Thessaloniki, Greece
e-mail: nikolaos@ideasforward.com

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_4

Breaking it down to bare essentials, who *you* are dictates *what* or *how* rules apply to you, be it written or otherwise. Rules still apply to you, even if your identity is not explicitly required for an action. For instance, in the case of traffic lights, who *you* are will be punished if you do not respect the rules. Generally, we exist and interact peacefully within our cities because they are governed by rules, customs, codes of conduct, which *we* must abide by. These are necessary building blocks of our cities, in order to function efficiently, interact productively, and ensure the stability required moving towards our future.

1.1 Identity in the Foreground

When we talk about identity we must tread lightly. A regulated identity, typically a national ID and/or passport, is the one that identifies us and allows us to sign contracts with others and interact with governmental and other services. But a person is not defined unidimensionally. Consider personal data like age, gender, religion, sexual orientation, ethnicity, personal health data, and others. Let us not forget social roles like motherhood, teacher, nurse, and others, or social group participation. Sociologists distinguish between personal and social identity and while it is our regulated identity that is officially used on paper, it is with these two that we interact at every other level. Add our digital self and the myriads of interactions that can be attested to it in the digital realm, be it using social networks or digital services, and one can see how identity can be powerful, yet fragile.

In a city environment, individuals' identities must be placed front and centre as misuse can cause ripples in the social fabric. It can traumatize the relationship between individuals and the city, affect the ties of social groups, give rise to animosity and resentment. Conversely, if treated with trust and respect, they can be a city's greatest springboard for urban catalyzation as suggested by Dr. Ali Cheshmehzangi (2020) in his book *Identity of Cities and City of Identities*. A city's identity is but the sum of its citizens' identities, thus the manner in which it incorporates them into its processes can make all the difference.

1.2 Rules of Engagement

At the base of every interaction there are rules dictating how outputs are defined by inputs, and even if these rules are not predefined, there are other rules that dictate how the outcome must be treated. Let us illustrate this with two simple examples, one evident and another less so. When you visit the bank to open an account the bank asks for your national ID and other details that (a) identify you as an entity and (b) allow the bank to run a check on you called Know Your Customer (KYC) check. When everything checks out, the bank gives you a contract, which you are expected to read carefully and then if you agree to the terms, sign it. The terms are the rules which dictate your relationship with the bank, your signature means that you have agreed and thus take full responsibility under the law to face the consequences if terms are not met. This relationship is transactional in nature, meaning that every time you transact with the bank or use your debit card, the transaction is logged and can be audited at any time. This is a purely rule-based relationship with predefined rules which most of the times will stop you from doing something outside its confines, e.g., overdraft, and in case this is not possible to audit you in the future.

The next example, however, is not that straight forward. You move to a nice neighbourhood with your family. Six months later a vibrant bar opens its doors, right next to your house. It respects quiet hours as dictated by law, but the rest of the day it is so loud and crowded that your family is affected, you cannot find a parking spot near your house and you start to feel resentful. To make matters worse, because of all the people coming in, the neighbourhood becomes a magnet for other similar businesses. Suddenly your choice to move there with your family becomes a nightmare and you are living in it. What rules govern this case? If the law is the only defining agent, then who serves the citizen's social identity?

Rules as previously stated can be explicit or implicit. With all aspects of identity being of equal importance, a city must uphold the rule of law but also address its citizen's needs. Rules are important not only at a daily transactional level, where it is easy to foresee and define, as well as at a forward-thinking level where a city can run simulations of itself growing, changing, and hopefully adapting.

2 Digitalising Interactions

A city is facilitated by a sum of rules, supporting processes, which are in place to serve groups of identities, which in turn are constituting the interacting entities. With cities growing in size, so did complexity, and processes became incumbent, inefficient and counterproductive. Fast forward, the internet era brought enablers that allowed us to digitize some of them, make them faster and more efficient. Consider logging in to online services with your email and password, handling your taxes, tracking assets in a logistics chain, e-commerce payments, health data, online applications, and so on.

This led to the first digital transformation, largely based on the identity aspect, leading several processes to transit from slow paper based bureaucratic processes to digitally enabled and streamlined ones. However, major problems surfaced like security flaws, personal data vulnerability, mistrust of third parties and their use of our data, accountability of participating entities and auditing transactional activities among others, all of which slowed down this digital transformation over the last couple of years.

2.1 Enter Blockchains

One of these problems is *double spending*. It affects the digital cash system and underpins the issue of a digital currency being spent twice. This is typically a fraudulent attempt taking advantage of security flaws of financial institutions handling the transaction of digital cash, leading to digital information being reproduced. Well, someone or a group of people under the alias Satoshi Nakamoto (2008), came up with a brilliant idea. It was suggested a peer-to-peer network that would allow transacting entities to cryptographically perform a transaction directly without a financial third party acting as a mediator. The underlying mechanism would have to rely on many distributed computer nodes agreeing (mining) on the validity of a transaction through a consensus mechanism called Proof of Work. This would in turn create a public chain logging transaction in an immutable sequence open for auditing by virtually anyone. The Bitcoin cryptocurrency was born.

The idea was revolutionary. A distributed network, dependent to no one entity? Allowing disparate parties to transact financially without a third party necessary to provide trust? Talk about security and democracy in one go, right? Well, not exactly. It turns out Bitcoin even though it became popular soon after and it is still in the spotlight with over $600 billion in market capitalisation, the single biggest money event in history, has major shortcomings. It has inherent scalability issues, is subject to fear of consolidation of major mining farms, does not fulfil certain characteristics to allow adoption by financial systems and at the end of the day it is only a digital cash system that cannot integrate with anything else.

You would think that would be the end of that, but you would be wrong. In 2013, Vitalik Buterin proposed another blockchain framework called Ethereum (Buterin 2013). The underlying concept was, and I quote from the Ethereum Org whitepaper:

> a blockchain with a built-in Turing-complete programming language, allowing anyone to write smart contracts and decentralized applications where they can create their own arbitrary rules for ownership, transaction formats and state transition functions

So, in 2015 Ethereum framework was launched publicly and it supported the following three components (a) a cryptocurrency call ETH, (b) smart contracts, or simply put pieces of code, residing in the nodes to govern rule-based transactions, and (c) the ability to write decentralized applications, i.e., small pieces of software that could be integrated anywhere and facilitate access. All these characteristics under the prism of a trustful environment supporting any group of disparate parties and based on unbreakable cryptography. *This* was a true game changer. Applying rules under complete cryptographic principles fuelled a myriad of applications, inspired thousands of researchers, and large enterprises trying to position themselves in this new field. Blockchains or better yet Distributed Ledger Technologies (DLTs) were considered to be of the same magnitude of importance to the internet as cloud computing was, when it first came out, if not more so.

2.2 Enterprise Blockchain Frameworks

Following the introduction of Bitcoin and Ethereum several other cryptocurrencies were born. It seemed that the democratization of digital cash was on the rise. But cash, be it digital cash or otherwise, is a means to an end and not all processes are cash bound. Which brings us back to the two cornerstone ideas of identity and rules governing processes. Besides that, public blockchains have an inherent characteristic that does not necessarily fit most of real-world scenarios, and that is they are public by design, i.e., open and permissionless allowing anyone to participate and access the transactions' log. This led to a new wave of blockchain frameworks being proposed, which were designed with corporate activities at the heart of them, suggesting permissioned only or hybrid architectures more suited to corporate and governmental environments.

The Linux Foundation started hosting one of the first open-source (OS) community called Hyperledger, which quickly grew to one of the largest communities in the world developing enterprise-grade blockchain frameworks, libraries, and tools. Hyperledger frameworks widely used for research as well as currently deployed in the market are Fabric proposed by Digital Asset and IBM, Sawtooth proposed by Intel, Iroha and others called general purpose frameworks based on the concept that actors interact with other actors or assets, whilst Hyperledger Indy is focused on decentralized identity management. To support these frameworks there are libraries like Hyperledger Aries and Ursa and tools like Hyperledger Explorer and Cello open to the community to quickly implement solutions. These frameworks were built and are still evolving with a cross industry focus in mind, with major applications currently in value chains, healthcare, and governmental fields.

Then you get frameworks like Ripple a brainchild of Ripple Labs focused on financial transactions like settlements and remittances. Corda span out of R3's research, open sourced soon after and initially focused on legal contracts but is currently fast taking its place as a cross industry framework. Quorum is another example that was designed by J.P.Morgan for the financial field, acquired by Consensys and subsequently open sourced and opened to other application fields.

These are some of the examples of blockchain frameworks out there, which is just a family of DLTs. The one common theme is the fact that they are all exploring the OS route to standardization and the market. The wider community is jumping at the opportunity to contribute to them and use them for market or research applications. Similarly, the big IT warehouses like IBM, Microsoft, Amazon Web Services, and others are exploiting the field with Blockchain-as-a-Service offerings and services. They are actively pulling their weight to roll-out large real-world applications working with financial institutions, value chain behemoths, the manufacturing industry, healthcare systems, and governments. Governments and institutions though latecomers in this field are also actively investing and exploring the opportunities that lie in adding this enabling technology in their arsenal.

2.3 Shortcomings and Limitations

As any technology in its infancy blockchain is not without shortcomings. In a comparison of open blockchains and DLTs conducted (Dinh et al. 2017) against several benchmarks the core issue is scalability. Fostering millions of transactions per second, whilst maintaining a secure-by-design architecture is a major roadblock. Granted not all applications have real-time significance but then again, a large enough number does. How long would waiting for an application approval be too much?

The main consensus mechanisms require high processing power making them resource intensive. Some frameworks, like Hyperledger Sawtooth, have adapted to allow pluggable on-the-go consensus algorithms, which means that as a blockchain based system scales, consensus algorithms can be adapted for speed and efficiency. This however may compromise the inherent characteristics of security. Another factor leading to security concerns is the 51% attack, which means that in a public blockchain whoever owns 51% of the nodes, basically owns the network. Interoperability is one other limitation, as most frameworks do not work well with each other, thus leading to siloed and disjointed implementations. Quantum resilience is also an issue for most blockchains, which are not ready to handle the advances of processing capabilities quantum technology will entail. Subsequently an already complex and resource intensive technology may need to become even more complex and resource intensive to compensate.

However, limitations are faced by any new technology out there. The effort put into overcoming them is directly proportional to (a) the application field and (b) the intrinsic value it has to offer. The first blockchain was proposed in 2008, and now, almost 13 years later, the momentum for adoption and diversification is stronger than it ever was.

3 Blockchains and the Real World

Cryptocurrencies were the first application of blockchain technology. Bitcoin, Ethereum, Cardano, Litecoin, and many others sprang out were crowdfunded, via so-called Initial Coin Offerings (ICOs), raising as much as Telegram's Open Network $1.7 billion, and some of them are still traded at crypto exchanges. But digital cash as explained is just one side of the proverbial coin. With enterprise-grade frameworks available, many more applications were implemented and are being researched, some of them already hitting the market supporting various industries and their processes.

3.1 Applications of Blockchain

Let us go through a few sectors where blockchain technology is leveraged or considered so far.

Logistics & Transport: In 2018, Maersk along IBM and GTD Solutions launched TradeLens. TradeLens is an open blockchain powered ecosystem for supply chain management as well as tracking of vehicles, containers, and assets. It allows disparate parties to share information across supply chains, reducing friction in global trade. Similarly, in 2019 the United Parcel Service (UPS) joint forces with Inxeption to launch Zippy, a B2B blockchain powered marketplace facilitating logistics and tracking for e-commerce businesses. And the list goes on with household names like British Airways, FedEx down to innovative startups.

Mobility: Mass mobility is also benefiting from the application of blockchain. Apart from major industry players like Accenture and Ernst & Young positioning themselves in the field, several startups are also bringing forward innovative solutions. The mobility space lends itself to a multitude of applications from ticketing and connected cars, to vehicle-to-infrastructure and dynamic insurance. Every year at MOVE in London, a mobility focused conference, a parade of implementations leveraging blockchain are presented to the public. Moreover, the Mobility Open Blockchain Initiative (MOBI) has formed to address the potential of blockchains in mobility, supported by major Automakers, Governments, Insurance companies, Academia, Cities, Startups, and many more.

Healthcare: There are but a few sectors where the application of blockchain can be more fitting than the healthcare sector. Health data security, sharing thereof, concept management are cornerstone in the field and there are many efforts attempting to address this. United States startup, SimplyVital Health is one such example creating blockchain powered open-source database for healthcare providers. Robomed, a Russian company partnered with Taipei Medical University Hospital to secure patients' medical records. At institutional level, the Centres for Disease Control and Prevention (CDC) in the USA is working with IBM on data collection solutions. European startups are not lagging behind in the race, but most importantly research through Horizon 2020 framework projects like *My Health My Data* are paving the way forward for patient data privacy, security as well as transferability across healthcare systems. Similarly, *Pharmaledger* a project sponsored by the Innovative Medicines Initiative (IMI) and the European Federation of Pharmaceutical Industries and Associations (EFPIA) under the Horizon 2020 framework programme is looking at blockchain scalability and governance to address major roadblocks of the pharma industry today like clinical trial data privacy and drug counterfeiting, among others.

Energy: Another promising sector primed for disruption is energy. The sector comprises of production, distribution, pricing, charging of Electric Vehicles (EVs), and other activities, all of which are fundamentally rule based. Apart from research projects active in demand response, microgrids and other subjects several companies are rolling out peer-2-peer energy sharing marketplaces like American startup LO3

Energy, whilst others like Israeli startup Greeneum, which incentivizes consumers to save energy, are looking to turning consumers to prosumers. Big companies are also active. IBM's Energy Blockchain Labs, for example, facilitates emissions trading, whilst Accenture is focused on procure-to-pay processes in the Oil and Gas market.

These are but a few examples of applications currently hitting the market, and we have not even touched the financial sector. The buzz is high, investments are monumental from both private capital firms and institutions. If nothing else, it proves that blockchain has the potential to facilitate viable solutions, although more must be accomplished for the technology to leave up to its promise.

3.2 Governments and Institutions

While businesses are taking care of business, institutions are also examining the use of blockchain technologies. As in any race, you get frontrunners and latecomers, and in the case of institutions like governmental and civil services you would expect the timescale of change to span across decades. However, as blockchains stand to offer several obvious benefits, some governments hurried up to experiment with the new technology even though it came out in a global recessionary period. Some of the top-level benefits that could be leveraged are

- Securing citizens' and businesses' data, with secure and cryptographic storage.
- Cutting down on bureaucracy and costs by automating processes.
- Limiting the opportunities for corruption and fraud, through traceability and immutability of records.
- Reinventing identity and support consent-based sharing of personal data.

Estonia was among the first nations to fully digitize processes like e-governance, e-tax, digital identities, e-voting, e-health, so it came as no surprise when as early as 2012 it designed the Keyless Signature Infrastructure (KSI) blockchain to serve its Justice Department, then expanded its use in Healthcare, Land Registry, and other services. Such was the success of the KSI blockchain that is currently used by NATO and the U.S. Department of Defence.

Estonia is not alone in this. The Netherlands has deployed the Pension Infrastructure which is a blockchain powered pension administration back-office. The Swedish Mapping, Cadaster and Land Registration Authority in 2016 launched a blockchain based land registry. Malta is putting educational credentials and transcripts on the blockchain, whilst also rolled out a favourable regulatory framework for blockchain businesses and cryptocurrencies. The list continues with more nations piloting blockchain solutions and implementing them day by day and they are not without support. The European Union, for example, has dedicated millions in research in the field, has setup the EU Blockchain Observatory and Forum, has launched the European Blockchain Services Infrastructure (EBSI) to provide institutional backing in the efforts to reach standardization, setting up distributed nodes across EU member states that will facilitate public services.

3.3 Cities in the Loop

Cities in this respect are smaller scale counterparts. Blockchain solutions built by governments affect cities in a big way as they have local civil services and part of national infrastructures. But they also have a closer connection to local businesses, communities, schools, healthcare facilities, citizens, and a number of other actors, not to mention the municipal infrastructure. Thus, research and the market are working hard on providing their input and solutions in identifying the right applications of blockchains within a smart city framework.

Research: Researchers are working hard on identifying the place for blockchain technologies in future smart cities. The underlying characteristics of the technology are thoroughly examined (Hakak et al. 2020) (Salha et al. 2019) to identify potential benefits that can be efficiently exploited. Other research work is focused on specific challenges, like how Internet of Things (IoT) implementations can be secured (Rahman et al. 2019) (Gong et al. 2019) at the edge, and how cities can extract infrastructural data in a trusted way.

Another strong focus of researchers is identifying exploitable synergies between enabling technologies. The convergence of blockchain technology and artificial intelligence in IoT networks (Singh et al. 2020) is one of the examples of research activity taking place. Sharma and Park (2018) examined the feasibility and performance of a hybrid software defined network with blockchain implementations. They researched how edge and core network elements can take advantage of these underlying enablers to support a wider smart city infrastructure.

Verticals are also researched in terms of how blockchain can provide and support value added services for a smart city environment. Treiblmaier et al. (2020) examined the various application fields that blockchain technologies could affect within a smart city framework. More specifically Rehman et al. (2020) addressed the vehicular infrastructure proposing a purpose-built network architecture, whilst Nam et al. (2019) focused on tourism, digging deeper into the potential of the technology and how it can disrupt the industry. Boulos et al. (2018) worked on the geospatial significance of securing healthcare systems and patient data, whilst other studies are focused on securing medical devices (Paliokas et al. 2019) addressing the connection between healthcare and a smart city framework.

The list of research activity is long, and it keeps expanding to multiple directions. Digging deeper into the technologies and frameworks available, addressing limitations as well as exploring promises, focusing on application areas and making the connection with a smart city environment more enticing and achievable.

Markets: Considering research backing and the number of people working in this field since 2012, especially following the launch of the Ethereum framework, market ready solutions soon surfaced addressing various applications fields within a city framework. Cities work with large IT powerhouses and integrators, like the prominent example of Dubai working with IBM and others, are trying out solutions developed internally like the Swiss city of Zug which decided to accept some municipal fees to be paid with cryptocurrencies, and they are also working closely with startups and SMEs developing innovative blockchain based solutions in a range of application scenarios.

Agora, for example, is offering a voting system powered by blockchain technology, which secures the voting process and is highly dependent on identity. Solutions like Agora, BallotChain, and others allow citizens to vote from anywhere, help cities reduce time and related costs and safeguard against compromise of local elections bringing real-time transparency to one of our most democratic processes. Penta is a research company creating solutions based on blockchain addressing supply chains, healthcare, fintech, and even smart city focused applications like fractional real estate ownership, paving the way to faster processes, rethinking ownership, and expanding access to real estate investments. Paradox Engineering is integrating blockchain to secure city scale IoT infrastructures supporting from parking spaces and streetlights to waste collection and management. The Suez Group has developed the CircularChain leveraging a more circular approach to waste management.

Other companies are looking more into future applications, like Greyp a mobility tech startup from Croatia, which partnered with slock.it, part of Blockchains, LLC, to develop automatic payments for mobility solutions. Similarly, Chorus Mobility is working on how blockchain can be the underlying enabler for future mobility under a peer-to-peer outlook where vehicles are communally owned, self-driven and interact with a smart infrastructure, whilst Iomob *Powers Open Mobility Marketplaces* in support of shared mobility business models. Other efforts are also gaining ground working on peer-to-peer energy exchange, like the example of the Brooklyn Microgrid supported by LO3 Energy. ClearTrace is focused on emissions' tracking, which could easily scale down from institutional level, to city, business even neighbourhood or home level and support a new emissions trading economy maybe fractional, and maybe lead to a greener future.

4 The Long Road Ahead

When one talks about smart cities, it is almost a utopian vision of a city able to understand its citizens' needs and provide for them at all levels. Adapt and safeguard its infrastructural integrity and efficiency, and work towards its own positioning in the grander scheme of things. This vision is based on identities and their fractal tangents, it is based on rules and governance, but most importantly it is based on the understanding of a better future.

Rules do not enforce communal mentality, they do not ensure that innovation happens, they cannot give rise to creativity and art. In fact, Bettencourt and West (2010) two theoretical physicists turned urban theorists, came up with a set of simple yet powerful mathematical equations that can fully characterize a city from its sewage system length to its number of violent crimes. Moreover, these equations bring the defining patterns that can lead to a city's future into the light. However, when superimposing that to historian Lewis Mumford's famous quote on the rise of the megalopolis described as *the last stage in the classical cycle of civilization*, to which West agrees, one is left with questions about how future can be defined and how sustainability can be achieved without compromising the unique characteristics that a make up a city.

Blockchain is just an enabler, and when integrated with powerful technologies like Artificial Intelligence, Big Data, IoT, and others, they will eventually make our cities … cleverer. Turning our cities into smart cities will take much more than that, though. It will require deeper understanding of how processes and data dictate results and reactions. How individuality and entrepreneurship affect a city in real time. How cities as part of global communities and economies can position themselves, be complementary to each other, support value exchange systems within as well as outside their limits. How they can tap into their internal potential to instigate uniqueness as well as create opportunities for a better future across the board.

Today we keep things simple, take small, calculated steps to achieve efficiency at scale, sustainability, build trust in sharing. But if our cities are to become smart, then they need to be able to define their *own* identity.

References

Bettencourt L, West G (2010) A unified theory of urban living. Nature, 467(7318):912-913

Boulos M.N.K., Wilson J.T., Clauson K.A. (2018) Geospatial blockchain: promises, challenges, and scenarios in health and healthcare. Int J Health Geogr 17, 25. doi:https://doi.org/10.1186/s129420180144x

Buterin V (2013) Ethereum Whitepaper. Available via Ethereum Org. https://ethereum.org/en/whitepaper/

Cheshmehzangi A (2020) Identity of Cities and City of Identities, Springer Nature

Dinh T.T.A., Wang J, Chen G et al (2017) Blockbench: A framework for analyzing private blockchains. In Proceedings of the 2017 ACM International Conference on Management of Data:1085-1100

Gong S, Tcydenova E, Jo J et al (2019) Blockchain-based secure device management framework for an internet of things network in a smart city. Sustainability, 11(14):3889

Hakak S, Khan W.Z., Gilkar G.A. et al (2020) Securing smart cities through blockchain technology: Architecture, requirements, and challenges. IEEE Network, 34(1):8-14

Nakamoto S (2008) Bitcoin: A peer-to-peer electronic cash system. Available via Bitcoin Org. https://bitcoin.org/bitcoin.pdf

Paliokas I, Tsoniotis N, Votis K et al (2019) A blockchain platform in connected medical-device environments: Trustworthy technology to guard against cyberthreats. IEEE Consumer Electronics Magazine, 8(4):50-55

Rahman M.A., Rashid M.M., Hossain M.S. et al (2019) Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city. IEEE Access, 7:18611-18621

Rehman M, Khan Z.A., Javed M.U. et al (2020) A blockchain based distributed vehicular network architecture for smart cities. In Workshops of the International Conference on Advanced Information Networking and Applications:320-331. Springer, Cham

Salha R.A., El-Hallaq M.A., Alastal A.I. (2019) Blockchain in smart cities: Exploring possibilities in terms of opportunities and challenges. Journal of Data Analysis and Information Processing, 7(3):118-139

Sharma P.K., Park J.H. (2018) Blockchain based hybrid network architecture for the smart city. Future Generation Computer Systems, 86:650-655

Singh S, Sharma P.K., Yoon B et al (2020) Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. Sustainable Cities and Society, 63:102364

Treiblmaier H, Rejeb A, Strebinger A (2020) Blockchain as a Driver for Smart City Development: Application Fields and a Comprehensive Research Agenda. Smart Cities, 3(3):853-872

A Cross-Domain Landscape of ICT Services in Smart Cities

Barbora Buhnova, Terezia Kazickova, Mouzhi Ge, Leonard Walletzky, Francesco Caputo, and Luca Carrubbo

1 Introduction

Smart cities have been a widely discussed topic over the last decade [24, 2]. Their potential to solve challenges related to the growth of urbanization, environmental issues, and worldwide trend of aging population are motivating factors for research in this area [3]. The research initiatives of smart cities focus on different areas, whose range of research topics is widespread, including urban planning, mobility and transportation, smart living and community, smart environment, emergency services, e-Health, and government, to name a few.

The smart city can be explained in terms of a complex of digital services exchanged by a network of actors interconnected in order to share knowledge, resources, competences, and capabilities to perform better. These digital services have been considered as one of the important components in a smart city since the digital service is capable of connecting service providers, users, infrastructures, and communities in a common ecosystem to support the value co-creation. However, the overall picture of the digital services included in each domain and their dependencies are still missing. There have been attempts to create overview models for smart cities, some of them are, however, quite complex, and the dependencies are

B. Buhnova (✉) · T. Kazickova · L. Walletzky
Faculty of Informatics, Masaryk University, Brno, Czech Republic
e-mail: buhnova@fi.muni.cz; kazickova@mail.muni.cz; qwalletz@fi.muni.cz

M. Ge
Deggendorf Institute of Technology, Deggendorf, Germany
e-mail: mouzhi.ge@th-deg.de

F. Caputo · L. Carrubbo
University of Naples Federico II, Naples, Italy
e-mail: francesco.caputo2@unina.it; lcarrubbo@unisa.it

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_5

usually missing in these models. We believe that this missing piece can bring new insights into the smart city research and also bring us to uncover new possibilities of interconnecting current services and providing value to the citizens.

In this paper, we therefore analyze smart city projects across the world to identify the emerging technologies and domain-specific services, which are then examined together with the relationships, influences, and dependencies among these services, as well as among different areas of smart cities. As the key contribution, we found an effective way of depicting the results of our analysis in a landscape model, interconnecting the same levels of abstraction for each domain of the smart city. We follow the five-layer model of Information and communications technology (ICT) services within smart cities proposed in [1], which is considered to provide layered outlook of the ICT services, in order to identify dependencies on different levels of smart services. Our concluded overview reveals many interesting findings about the similarities and differences across the studied domains, which can facilitate the development of a holistic smart city platform that would effectively integrate these services.

The remainder of the paper is organized as follows. Section 2 reviews the related works and highlights our contributions. Section 3 describes the layered model of the smart city. Based on this model, Sect. 4 investigates the technologies and services in different domains. Section 5 further discusses and interprets the research results while identifying various similarities and differences across the domains. Section 6 then presents reflections on integrated approaches in existing smart cities, and Sect. 7 discusses the implications from the managerial and value creation perspective. Finally, Sect. 8 concludes the paper and outlines future works.

2 Related Work

Over the time, the city management has attracted the interest of both researchers and practitioners from different research streams [4, 55, 41, 20, 53]. Several contributions have discussed how structural elements of cities can be combined [73, 74, 75], to the role of citizens and organizations in ensuring city viability [76, 77, 78], and how one can react to improve the quality of life in the city [79, 80, 81]. In the last decades, the studies about city management have been strongly influenced by the potential contributions provided by the advancements in knowledge rooted in computer science [82, 83, 84]. ICT have progressively become essential part of the frameworks on which city management studies are built [85, 86, 87]. As a consequence of the evolution of knowledge, ICT has progressively shifted to be considered a support for a better definition of city management approaches [88] and become central enabler to influence a radical change in city development [89, 90, 91]. The more recent evidence of this change is usually referred as the "smart city" label [92, 93].

In the smart city research, more than 70,000 contributions can be found in Google Scholar but a widely accepted definition seems to be still missing [89].

With the aim to bridge this gap, some researchers have focused on the role of technologies in supporting citizens' participation to city management [94, 95, 96], while other contributors have analyzed the ways through which the city evolves as a consequence of a more efficient use of available data and resources [97, 98, 99].

Among the multiple definitions provided about the concept of smart city, Hollands [100] defines the smart city as an "urban labeling" phenomenon that does not identify a real concept but only a change in perspective on the ways in which the city is approached. From a different perspective, Bakici [101] refers to the smart city as a domain that involves "implementation and deployment of information and communication technology infrastructures to support social and urban growth through improving the economy, citizens" involvement and governmental efficiency" (p. 138). Moreover, Paskaleva [102] defines the smart city as "one that takes advantages of the opportunities offered by ICT in increasing local prosperity and competitiveness—an approach that implies integrated urban development involving multi-actor, multi-sector and multi-level perspectives" (p. 411) while Hernández-Muñoz [103] state that "smart cities can represent an extraordinary rich ecosystem to promote the generation of massive deployments of city-scale applications and services for a large number of activity sectors" (p. 459). Based on our review, previous contributions can be classified in three categories, as summarized in Table 1: technical, organizational, and social.

As shown in Table 1, smart cities represent a wide, multi-disciplinary, and multi-dimensional area of research. Building upon the multiple interpretative framework developed for describing the smart city, several models have been proposed for

Table 1 An overview of contributions about smart city domain

Perspective	Brief description	Literature contributions
Technical	From technical point of view, the smart city is analyzed as complexity of technologies combined for improving the general efficiency of city management in terms of improving in quality of city services, supporting citizens interaction and participation, and reduction of environmental impact among the others.	Chourabi et al. [104]; Hancke [105]; Lombardi et al. [106]; Deakin [107]; Mulligan and Olsson [108]
Organizational	From organizational point of view, the smart city is considered a multi- and trans-disciplinary domain able to increase the efficiency in linking institutions, services, citizens, resources, and processes.	Nam and Pardo [109], Anthopoulos and Vakali [110]; Lee et al. [111]; Anthopoulos and Fitsilis [112]
Social	From social point of view, the smart city is investigated as framework through which one can increase the quality of life of citizens from multiple perspectives such as relationality, access to city services, and increase in citizens engagement among others.	Leydesdorff and Deakin [113]; Vanolo [114]; Saviano et al. [115, 116]; Caputo et al. [117, 118]

analyzing the complex smart city landscape [24]. From a general point of view, smart cities are typically perceived as systems of systems (SoS) [119]. Each system represents a group of services related to a certain area of interest such as urban planning, public lighting, energy, mobility, environment, emergency, e-Health, and e-government [120]. Although these models provided relevant advancements in knowledge, they seem to be unable to describe the interconnection and mutual influences among city services [121, 118]. Specifically, recognizing previous managerial and theoretical contributions about smart city, the following weaknesses in depicted models can be identified as building points on which one shall act for defining a more efficient view for smart cities:

ICT as a Service, not an Underlying Layer The models that provide a holistic view of smart city landscape often consider ICT just as one part of smart city [109, 119]. In contrast to this view, ICT is a fundamental part of smart city which underlies and interconnects all the city services.

Abstraction from the Shared Infrastructure Several models approach city services individually [122, 123]. They assume that all the services use the same infrastructure. Once the shared infrastructure becomes congested, it can cause problems to city services performance and availability.

Abstraction from Citizen-Relevant Services Models that analyze city services are often more concerned about detailed specifications of the infrastructure while disregarding the citizen-level services. For instance, in case of the smart grid, the focus is often on the technical infrastructure, smart grid infrastructure, or smart grid itself [124], rather than the services built upon these.

In order to tackle these weaknesses, this paper recognizes the need for defining a shared approach that can provide a holistic service overview within the smart city, where ICT is considered a necessary step for developing an interconnected layered outlook of smart city structure.

3 Layered View of Smart Services

To unify the structure of smart city services and to build a basis for their interconnected holistic description, we introduce the concept of a layered view of the smart city [1]. The approach is based on service value proposition where the structure emerges automatically from the ordering of services according to their purpose.

In this model, five layers of services are identified where the layer one is proposing the value to the final user such as city citizen. Services from lower layers are providing their functionality to services from the upper level. These five layers are: (1) *smart features*—complex services, offering high perceived value to the city citizens. The value proposition depends on a particular configuration of services from the lower levels (e.g., mobility); (2) *smart services*—complex services that are using other (more simple) services. Their value proposition is aimed at smart features, even the possibility to use them directly is not excluded, but very limited (e.g., traffic control); (3) *support services*—simple services with predefined API that you can use to obtain particular information (e.g., vehicle of public transport position check); (4) *software*—layer that contains all basic software systems that are used to collect, store, process, or control the data; (5) *hardware*—layer of basic devices to get the data, e.g., sensors, activators, servers, and networks. This approach allows us to model the structure of the services across different domains and it identifies the smart service system in which the value can be perceived, diffused, and co-created on different layers of the service structure.

4 Smart Services Across Different Domains

This section presents the detailed layered models for eight smart city services, i.e., urban planning, energy, mobility, public lightning, environment, emergency, e-Health, and e-Government, which are being recognized as the key representatives of smart city services.

4.1 Urban Planning

Building a smart city needs to take many factors into consideration. To create a city that can adapt citizens needs and demands as well as changing technologies, it is necessary to approach the smart city design in a holistic way. Urban planning aims at connecting all the other parts of the city into one interconnected functional entity. Therefore, urban planning can be defined as "a technical and political process concerned with the control of land use and design of urban environments, which can benefit from trace data, analysis and mining" [30, 31]. We consider that urban planning creates prerequisites for all other services within the smart city and is responsible for deciding how many infrastructures the city needs, how they should be distributed and whether the infrastructure is sufficient for the needs of citizens. We can therefore state that "an efficient urban planning can, without any doubts, improve the quality of the life of all the citizens" [24].

Important tools for efficient urban planning are data tracing, data mining and analysis. There are open datasets as well as preserved datasets from the enterprises. Currently, due to crowdsourcing initiatives, when citizens play a role of sensor (usually by means of their smart phones), open data contain a wide range of data about location of citizens. On the other hand, preserved data from telecommunication companies also provide valuable insights into mobility across the city. Additionally, there are plenty of datasets regarding demographic and geographic information provided by the city. These datasets represent a valuable source for data mining and analysis and can support efficient urban planning [31]. As plotted in

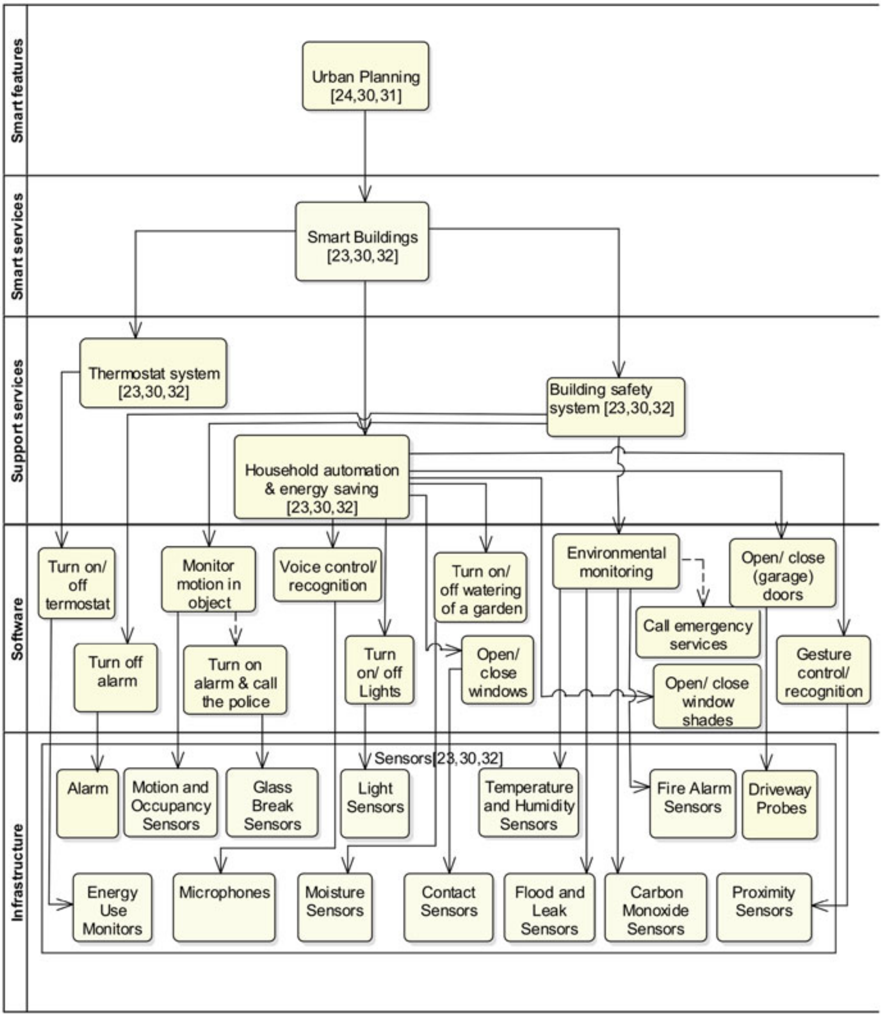


Fig. 1 Smart city services: urban planning

Fig. 1, urban planning includes a wide range of services. One of the most significant groups of services is smart buildings. Smart buildings are defined as the buildings with intelligent features such as ability to measure, store, and analyze data from the environment, namely for the purpose of household automation [13, 8], energy savings [23], or building safety [32].

4.2 Smart Energy

One of the fundamental elements in smart cities is the optimization of energy use within the entire community, which aims at achieving eco-friendly lifestyle with high quality of living [27, 34, 7]. The crucial role in this process is operated by ICT empowered electricity grid, which is known as the smart grid. In the smart grid, the ICT infrastructure improves efficient use of the physical infrastructure, providing the capacity to safely integrate more renewable energy sources and smart devices, delivering power more efficiently in secure and reliable way through new control and monitoring capabilities. By using automatic grid reconfiguration, the city can prevent or restore outages and enable consumers to have greater control over their consumption of electricity [35, 36, 37].

Several concepts of smart grid development and implementation have been introduced [17]. However, a complete picture of the smart grid ICT architecture and its existing design alternatives is still missing. When building the view in Fig. 2, we studied smart grid implementation from EU and USA. A valuable input for our study was the DISCERN project [18], which resulted into the generally accepted smart grid Reference Architecture (SGAM) model, as well as a use-case approach to smart grid development, where the use cases help to determine the key performance indicator (KPI) for the smart grid architecture [10]. Another project, called Grid4EU, presents six demo architectures. All of them were created with regard to SGAM. One of the demo architectures was proposed by a Czech energy distribution company [18]. This proposed solution focuses on automatic operation within the grid. It is supported by remote-control devices and connections that enable fast communication via regional dispatcher system. Another useful information source about smart grid architecture development in the USA was provided by a study of the California State University of Sacramento [38]. Although its main focus was cyber security and vulnerability of smart grid architecture, this study also brings information about the architecture itself and its stage of development in the USA. Finally, the description of the hardware part of the smart grid infrastructure can be found in [21], where the purpose of all hardware components in Fig. 2 is described.

4.3 Smart Mobility

Mobility is one of the core services in smart cities. Smart mobility is closely related to smart energy, especially electric mobility, which is directly dependent on smart grid infrastructure [28]. This area of smart city services experiences the rise of interest in recent years. It is expected that the smart mobility will be one of the main priorities of smart cities in the following years [48]. The main services included in smart mobility are public transport, electric mobility, and transportation management (see Fig. 3). Additionally, smart mobility includes support services

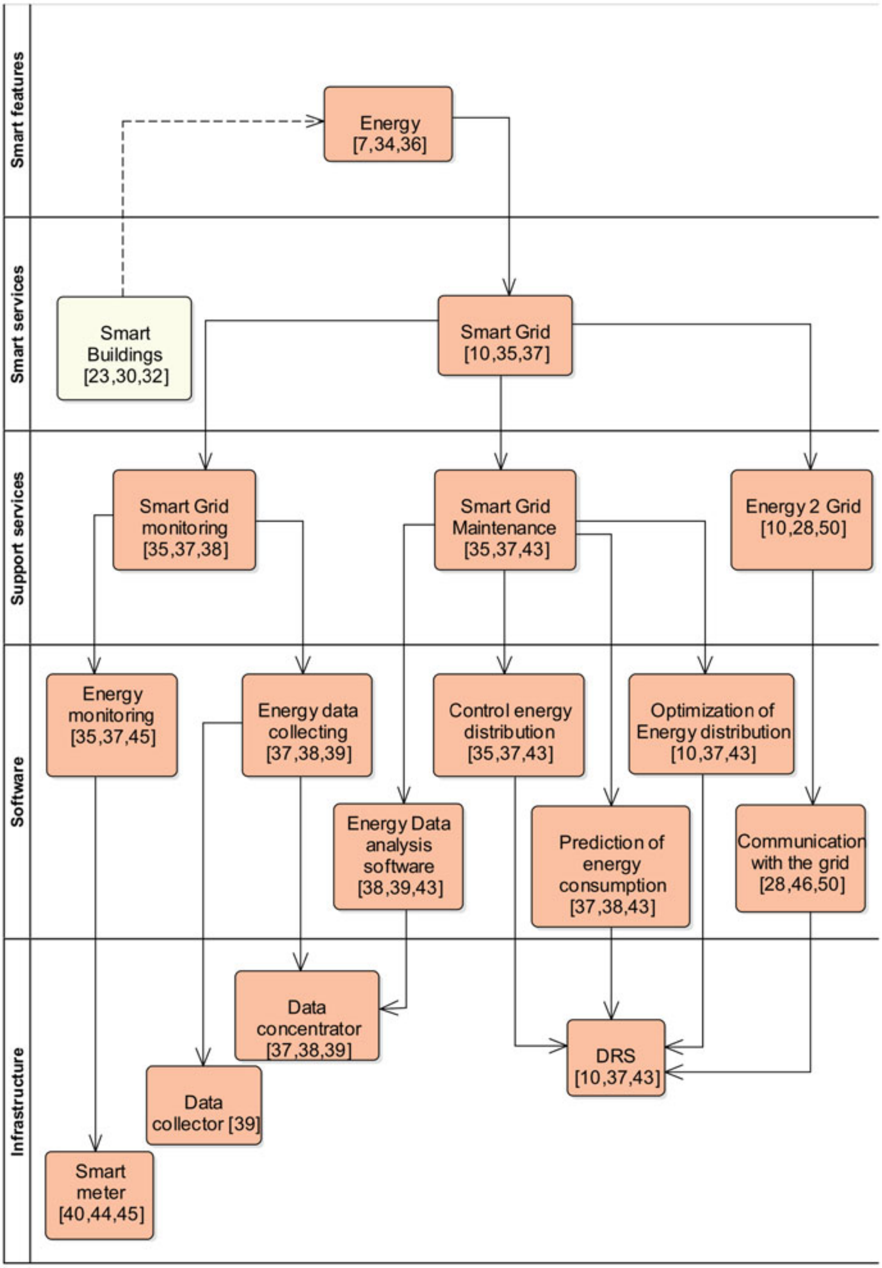


Fig. 2 Smart city services: smart energy

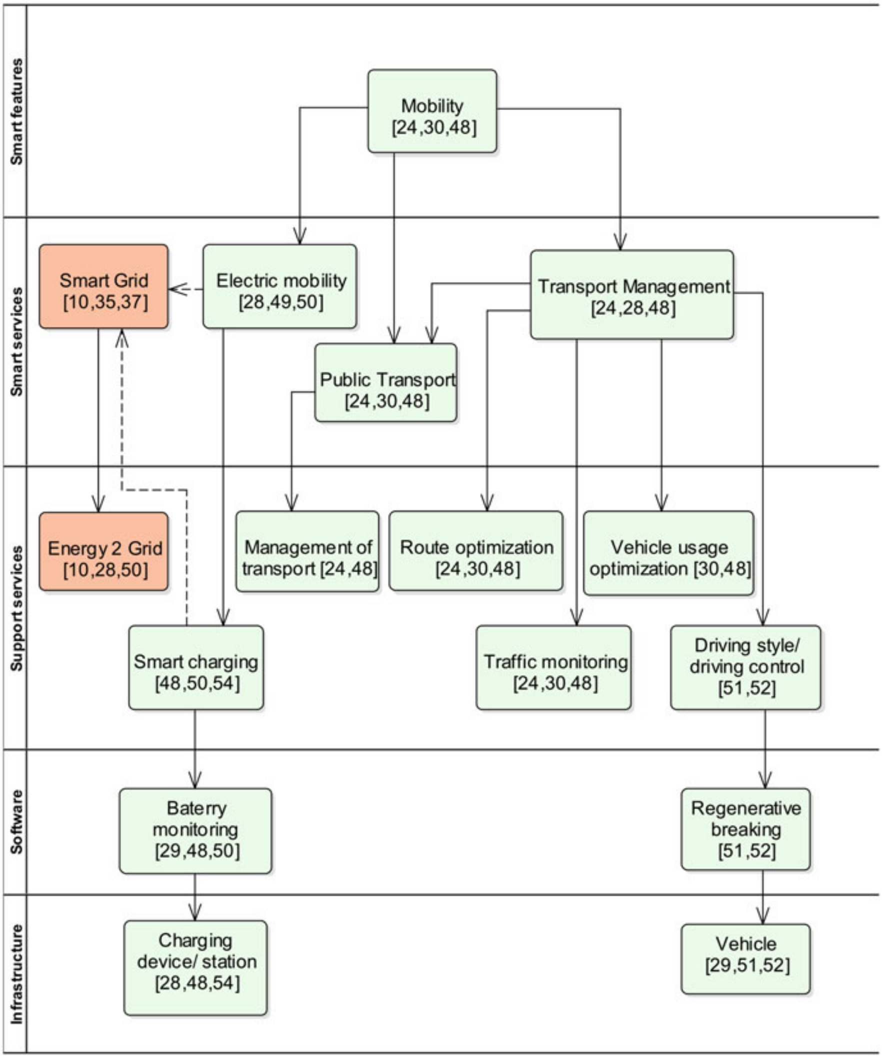


Fig. 3 Smart city services: smart mobility

like route optimization, traffic monitoring, vehicle usage optimization, and driving style control. These services are supported by software components which directly interact with the hardware elements on the foundational level.

Public Transport This service aims at providing citizens with reliable service that helps them to comfortably and efficiently travel across the city. Public transport makes use of the route optimization support service that makes the transportation more efficient. Another useful service is vehicle usage optimization, which takes care of distributing the vehicles on the public transport routes in an optimal way.

Electric Mobility As far as we found, this is the most dynamically developing aspect of smart mobility, focusing on creating conditions for plug-in vehicles within the city. That means that the city needs to create the infrastructure of charging station as well as smart grid infrastructure as a basic requirement for implementing electric mobility. Therefore, the electric mobility is closely interconnected with smart energy and smart grid. So far, the relationship is only one-directional, which means that the communication flows only from the smart grid to the electric vehicle. However, there are proposed concepts for vehicles-to-grid (V2G), which is plug-in electric drive vehicles (EDVs) with power-management and other controls that allow them to respond to external commands sent by power-grid operators or their affiliates, when parked and plugged-in to the grid [49].

Transport Management Systems This aspect of mobility includes support services such as route optimization, traffic monitoring, driving style control and vehicle usage control [30]. Traffic monitoring is one of the crucial support services in transport management. It provides us with the data gained from real-time traffic situation, based on which we can adjust our services in a timely way. Furthermore, a vehicle is a practical role of a sensor [30, 25], which collects the data about its environment. Based on this data, we can identify the places of traffic jams and apply this knowledge to route optimization. Moreover, information about the smart mobility gained in this way can be of high value for the people in other areas, for example, for emergency or security [30].

4.4 Smart Public Lighting

Another area closely related to urban planning is public lighting. It is estimated that public lighting is responsible for 5 to 60% of municipal electric bill [26]. Public lighting plays an important role for quality of life in the city. Firstly, public lighting ensures the safety of pedestrians on the streets and overall sense of safety in the city. Secondly, it is a requirement to lively city atmosphere, since it creates welcoming atmosphere for tourists and businesses [26].

One of the key tasks of smart city is to provide public lighting service with regard to energy efficiency and infrastructure. Layered structure of this service is represented in Fig. 4. Current public lighting services either support scheduling of lighting periods beforehand, or rely on an in-build sensor which turns on/off the public lighting automatically based on the amount of daylight. To control the mechanism of turning on/off the public lighting, there is a support control system available [11]. The mechanism for turning on/off the lighting can be optimized based on data analysis collected by sensors built in the public street device. The public light device usually includes a failure detector, which sends the information about the problem and the device can be quickly repaired. Public lighting systems support

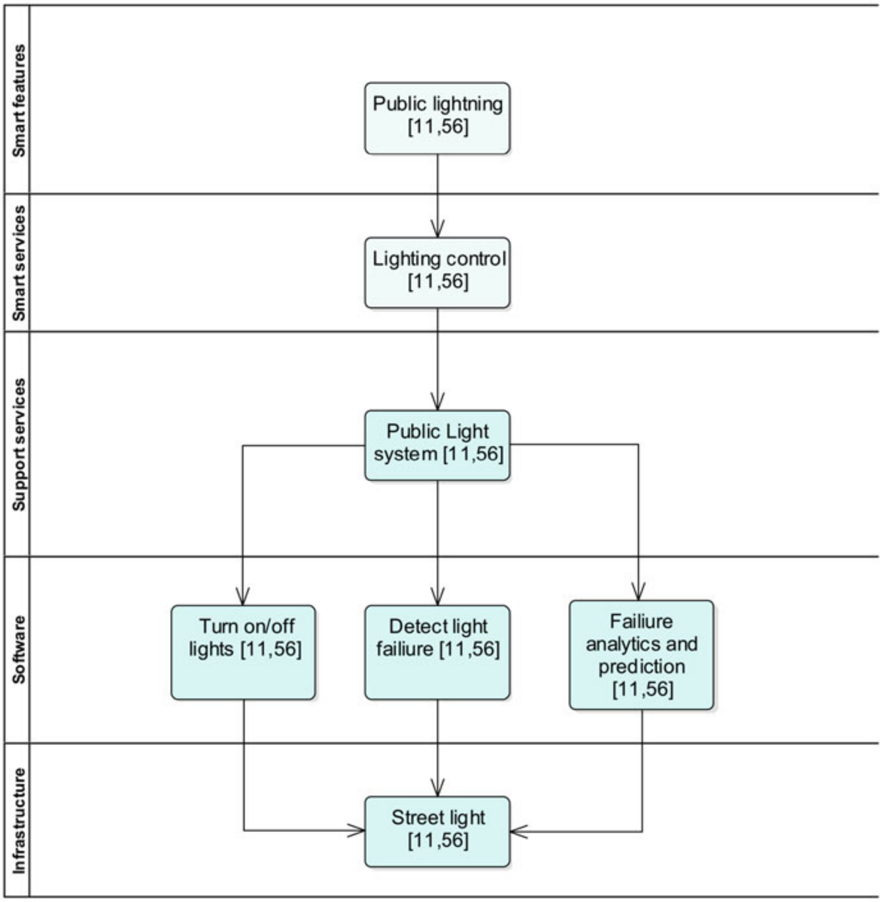


Fig. 4 Smart city services: public lighting

also a prediction of next failure and scheduling the regular control of the street light device accordingly. This prediction can be realized based on the data about previous failures and repairs.

4.5 Smart Environment

With growing urbanization, development of industry and transportation, the environmental problems become even more important. To provide a high-quality living condition, smart cities need to implement services that help to create a sustainable city. The main tasks of smart cities in regard of sustainability (see Fig. 5) are discussed in this section.

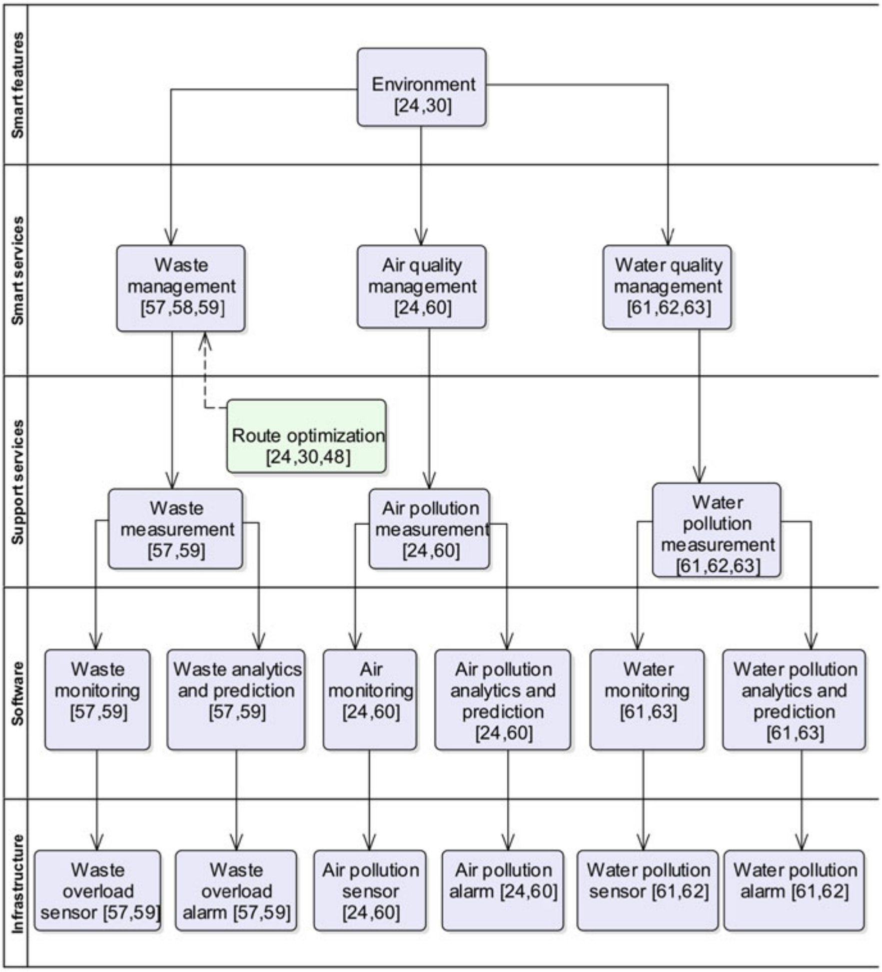


Fig. 5 Smart city services: smart environment

Waste Management Waste management means generation, prevention, characterization, monitoring, treatment, handling, and reusing residual disposition of solid wastes [22]. In the smart city, the waste is often recycled and recovered to make organic manure and generate power that can be consumed in smart city itself [22]. The goal of smart waste management is to implement independent system (with minimum human interaction) for handling the garbage in a way that reduces garbage and saves time, environment, energy, and money. Demands of smart waste management can be divided into three main categories: (1) Route optimization for garbage trucks [22, 58, 9]; (2) Monitoring and prediction of trash level based on sensors for remote trash level reading to optimize garbage collection as well as to optimize garbage bin placement [9]; (3) Smart waste collecting to automate waste transportation for waste collection center by means of underground suction pipes at the speed of 100–140 km/h [22].

Air Quality Management Monitoring of air quality plays an important role in urban areas with high density and increased pollution. In case of potential risks regarding air quality, the citizens can be informed about the situation and make further reactions accordingly. One of the issues by air monitoring is how to obtain real-time and accurate data about the situation in different areas of the city. This problem can be solved by placing the sensors on a vehicle driving across the city. There have been also attempts to monitor air quality by placing a sensor on bicycles. One typical solution is to use the vehicles of public transport, garbage trucks, taxis, or street sweeper vehicles. The gained data can be further analyzed by mining the data patters for improvement, or worsening of the air quality. Based on these patterns, the air quality can be predicted in advance and citizens could be warned [6].

Water Quality Management To control water quality, smart sensors can be implemented to gain the real-time data [19]. These data can be further analyzed. In case of risk, the citizens can be notified. This smart management usually involves sophisticated method of analysis and prediction. One efficient way of communication and warning is also available to the citizen, e.g., with smart phone applications. Combination of sensor information with crowdsourced information (delivered by citizens themselves) can bring valuable insights in the water quality and can help to protect citizen's heath in a more efficient manner [33]. Also, smart water quality management includes waste water management. Composition of waste water influences the resources of the water [33, 12]. For example, by heavy rains the sewer water can get to the rivers [12]. Therefore, it is advisable for the city to test the waste water for dangerous chemicals and to manage the flows of the waste water to save the water sources from pollution and people from risking their health.

4.6 Emergency

The goal of the smart emergency of a city is to ensure the safety of the citizens. The crucial parts of the emergency services are identifying the emergency and notify the related emergency service. Additionally, based on the gained data about emergency situations, we can predict their occurrence in the future.

Identify and Predict Emergency Situations To identify and predict emergency situations, it is necessary that smart city supports interoperability and communication of different systems. For smart emergency, an especially interesting part is the data from smart mobility.

Emergency Services—Ambulance, Firefighters, Police The main parts of emergency include ambulance, fire and rescue service and police (see Fig. 6). The most important factor for emergency services is time. This is the main objective when developing smart emergency systems [48]. The main reason for time delays in

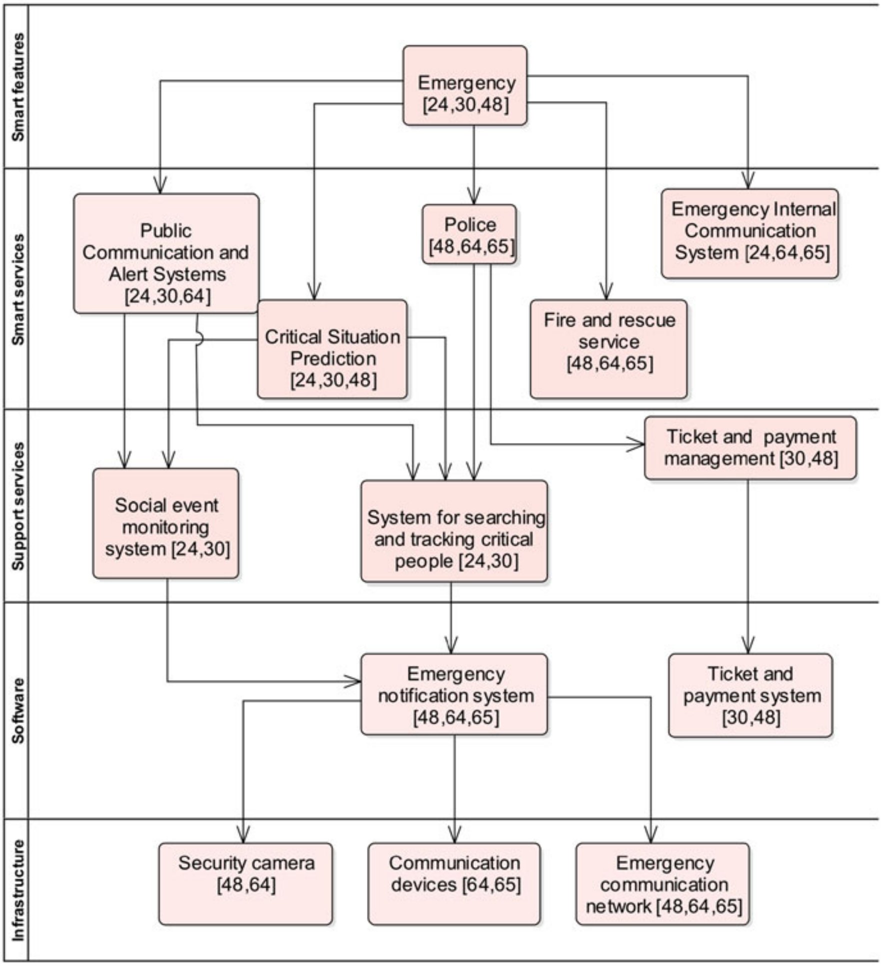


Fig. 6 Smart city services: emergency

urban areas is the traffic congestion or an obstacle on the route. In order to get the emergency services to the place of accident as soon as possible, it is necessary for smart emergency systems to communicate with smart mobility systems. Firstly, based on traffic information from smart mobility, the route for emergency vehicles can be dynamically optimized.

Secondly, by interoperability of smart emergency with smart mobility, the exchange of data between the emergency vehicles and traffic lights is possible. The communication with traffic lights combined with geo-fencing approach enables the emergency vehicle to cross the traffic light on green light at (almost) all situations. Geo-fencing is a feature in a software program that defines geographical borders [14] and enables us to identify the location of the vehicle and send an alert when vehicle leaves the borders (e.g., important for police vehicles when tracking the vehicle theft). Every emergency car has its own ID. When the car approaches the traffic light, and the traffic light identifies it as an emergency vehicle, the lights remain green until the vehicles passes the intersection [14].

Thirdly, another possibility to improve arriving times of emergency services as well as safety on the road is vehicle-to-vehicle (V2V) communication. This emergency vehicle warning system is built on the ability of emergency vehicles to transmit radio signals and detailed route maps to other vehicles and signals in their path for those vehicles and people to take appropriate and timely actions [15]. When the emergency vehicle is an ambulance, the hospital is notified about the location of the vehicle and its estimated arrival time [14]. The quicker arrival to the hospital as well as better informed hospital staff at the arrival of the ambulance, the higher is the survival rate of the victims of accidents or disasters.

Internal Emergency and Communication Systems On occurrence of traffic accident or disaster, man-made or natural, the cooperation of ambulance, firefighters, and police is needed. In emergency situations it is important to have access to the information and to have the possibility to communicate and be coordinated. This goal is fulfilled by internal emergency communication systems.

4.7 E-Health

E-heath refers to the broad range of information technologies, systems, and tools deployed in healthcare applications, including software systems for health data management and processing/analysis, computer assisted clinical process and decision support systems, and systems interoperability components (including software supporting healthcare messaging, telehealth, and medical devices [66]. e-Health provides health services by means of ICT. Implementation of e-Health systems brings improvement in healthcare, mainly in communication between patients and doctors, but also among the doctors participating in the same case (see Fig. 7). Furthermore, it facilitates decision making when choosing the treatment, eases health problems prediction, and it can also provide additional information needed when providing healthcare [67]. Currently, further new services related to remote patient monitoring and assistance are being developed [5].

Core system of the e-Health services is an electronic health records (EHR) system [68, 69]. EHR can be defined as digitally stored healthcare information about individual's lifetime with the purpose of supporting continuity of care, education and research, and ensuring confidentiality at all times [69]. EHR system enables the patients to comfortably access their health records as well as prescriptions. Moreover, it enables the doctors treating the same patient, specialists in different areas, to access all the information about the patient health history and treatment. That supports better sharing of information and, therefore, better decision making

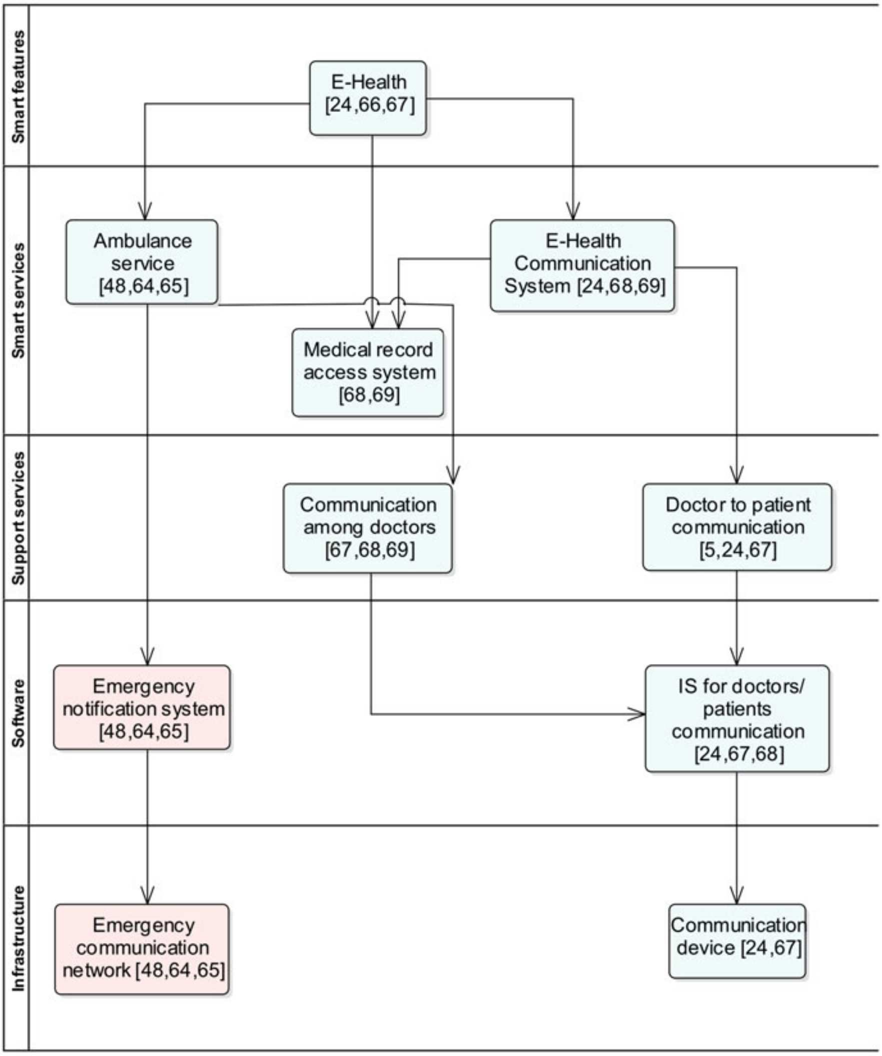


Fig. 7 Smart city services: e-health

in treatment as well as higher efficiency of communication [66]. Additionally, e-Health provides the appointment system [5] where patients can create reservation for medical examinations more efficiently.

e-Health services use highly sensitive, private information, therefore the expectations on such systems are very high, especially regarding security requirements, and the management of access to the information in two aspects: to ensure the access to the data for the dedicated person, and secondly, to ensure that nobody unauthorized will access the information [66, 68, 69]. Apart from security, there are additional requirements for e-Health systems that all e-Health systems must meet [66]. An important one is the interoperability with other systems, either other healthcare systems or managerial information systems (e.g., information system of insurance companies).

e-Health services include useful information not only for individuals but also for the general public. The data could be used for data analysis and data mining to find patterns and prediction [24]. For data mining and tracing, personal information can be filtered out. The possible benefits of data mining and tracing regarding e-Health data are, for instance, the monitoring of patient behavior, which can indicate health problems early and reduce future health issues [30].

4.8 E-Government

E-government means government that employs ICT in activities addressing the public sector [70]. The purpose of e-government is to provide easy access to information and mediate communication between the government and the citizens (G2C), between the government and other government institutions (G2G) and between the government and the businesses (G2B), while making it more democratic, transparent, and efficient [71].

The e-government concept is still being in development and its maturity varies in each city. The following main research areas of e-government can be identified [72].

Citizen Participation and Engagement E-government needs to provide quality services and to engage the citizens. By citizen engagement, we mean extensive use of technology by citizens to interact with the governments [71]. Citizen engagement and participation can drive government decision making, although citizens need access to more relevant and complex information in order to provide valuable insight into decisions of the government. This requires the government to publish relevant data openly in appropriate form for humans or machine reading and interpretation. Providing open data to citizens seems to improve their engagement with e-government. Social networks also seem to play their role in citizen participation. Thanks to social media, the community is formed, helping to create a community that the people are more interested in participating in [70].

Service Co-design Citizens participating in the development of e-government in their city can provide feedback, suggestions for improvement or data for analysis, which can be used during the design of services [42].

Standardization Standardization is necessary for providing effective and high-quality service. It is necessary for the interoperability of systems and interchanging of information between the systems of different areas of smart city as well as between the different systems within the e-government.

The main services provided by the e-government include sharing the important documents, communication with citizens, businesses and institutions, and informing

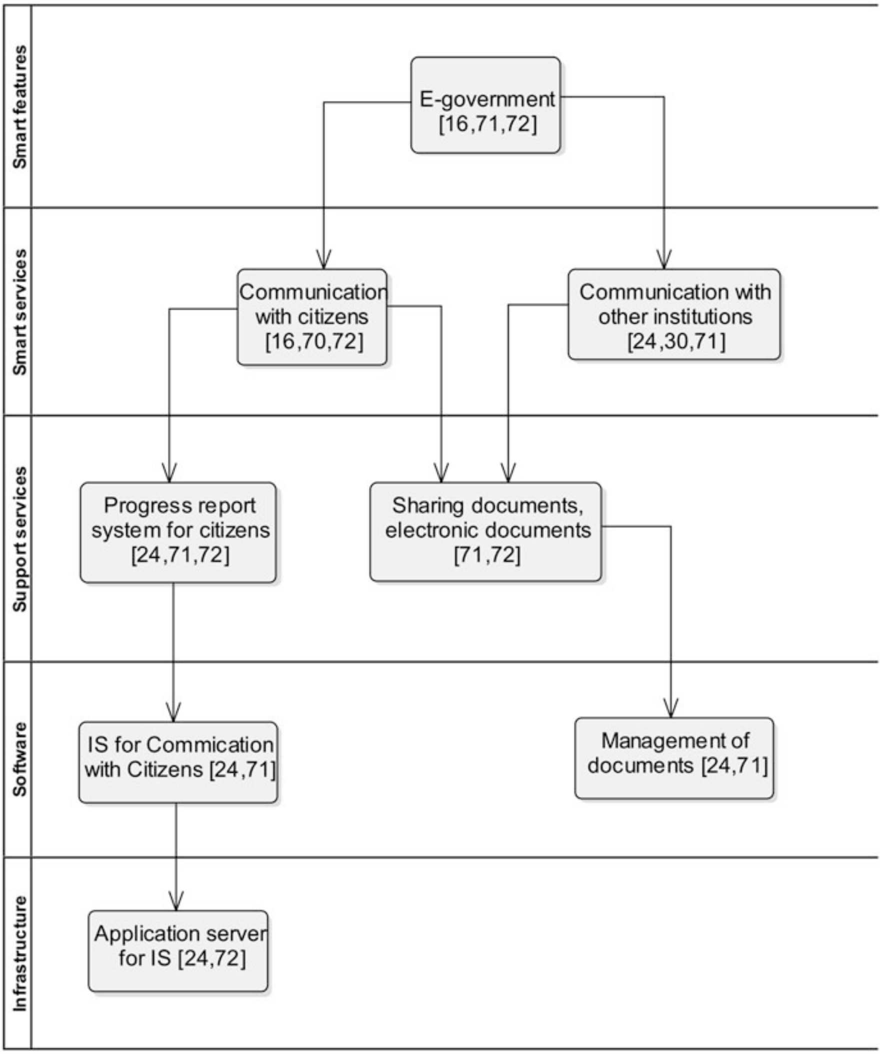


Fig. 8 Smart city services: e-government

the citizens about the progress of fulfilling the tasks of the institution (see Fig. 8). These services are typically supported via an information portal, which enables the communication from citizens to institution and vice versa, as well as document sharing.

5 Discussion

The key benefit of plotting the services onto a unified layered template is the possibility to compare their similarities and differences, and identify their interactions. An especially important benefit are the similarities of the services as it can motivate the synchronization and inter-operation during the real design and deployment of the services. The key findings from the analysis of the smart city services can be structured into the three main areas:

1. Similarities of infrastructure on the lowest level.
2. Similarities of services and software on the middle level.
3. Importance of alignment of smart services into the unified environment, which provides citizens with access to the individual smart services.

5.1 Similarities of Infrastructures

We found out that the elements of the infrastructure level are usually in the role of sensor or an actuator. As such, there are similar problems and tasks regarding their use, maintenance, and security (e.g., firmware updates). By looking at the wider context of these findings we can categorize smart features into three categories:

Smart Features Represented Across all Levels, Therefore, also on the Lowest Level This category includes urban planning, energy, mobility, public lighting, environment. The smart features in this category are closely related to Internet of Things (IoT) because they directly manage the sensors and actuators on the lowest level.

Smart Features Partly Represented on the Lowest Level This category includes emergency and e-Health. Even though emergency and e-Health are represented across all the levels (See Figs. 6 and 7), it is different than the previous category. The functionalities using the sensors and actuators are not widely used, and in most cases they only use the infrastructure for their emergency communication network.

Smart Features that Are not Represented at the Lowest Level This category includes e-government. The purpose of e-government is mostly to provide information that is not related to functions using sensors and actuators of the lowest level.

5.2 Similarities of Services and Software

Similarities at the level of services and software brought us to considerations of possible synergies in future development of the respected smart features. For example, as several smart features are using monitoring as well as controlling, it could be interesting to consider creating a unified infrastructure that would enable better decision making. Based on our analysis, we have decided to divide smart features according to the similarity of their functions at the second level as follows:

Urban Planning, Public Lighting, Energy Unified system could promote more efficient energy use. By collecting data about the number of people passing by the street light, we could optimize the placement of the street light, the time when it is lit as well as energy type which is used by public lighting. By unified system, many of these could be done dynamically. Similar system is already used in traffic lights. The same mechanisms could be probably used by public lights too.

Mobility, Environment Already, there have been projects which tried to connect mobility and environment. Specifically, smart vehicles driving around the city can have in-build sensors for air pollution [6]. Similar system could be used by water monitoring where the sensors could be placed on the ships.

Emergency, Mobility Unified system could help emergency services to get to accidents quicker. emergency systems could use the information from mobility systems.

Emergency, e-Health Unified systems could help us to predict risk patients (e.g., for heart attacks) based on their health records from e-Health systems. emergency services could have a prepared database of such people, so that in case of emergency they would be better prepared (pre-computed quickest ways to the home of the patient, detailed diagnosis, etc.) which could improve the survival rates in such cases.

E-Health, e-government e-Health uses sophisticated communication system for communication between patients and doctors. Both, e-Health as well as E-government focus on high security regarding personal information. Therefore, the communication systems for e-Health and E-government could be developed together by using similar mechanisms.

5.3 Unified Environment for Access to Smart Services

At the top level (smart features) we find it highly useful to create a unified environment (application or web portal) which would enable citizens to manage all provided services. Providing such a unified environment could help citizens to use smart services and participate in smart city. Additionally, it is not rare that people are hesitant about the level of technology integration in smart cities and smart services. Such an application which would enable access to all the services from a smart phone could help citizens to get involved and make use of the services.

The inputs for such an application are reflected in the overview diagram (see Fig. 10) which provides simplified overview of citizen and ICT services landscape within smart city. As distinct to previous diagrams (see Fig. 9), we have divided the

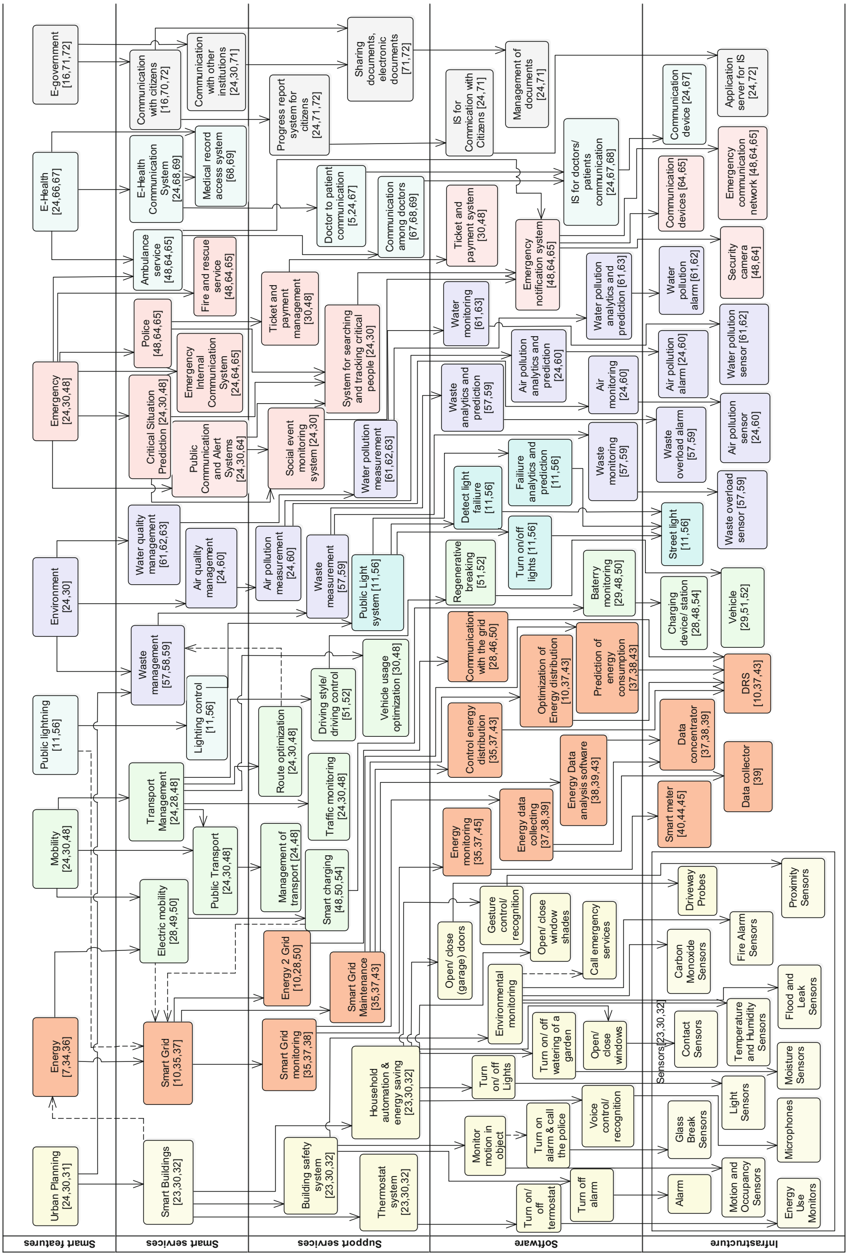


Fig. 9 ICT services of the smart city

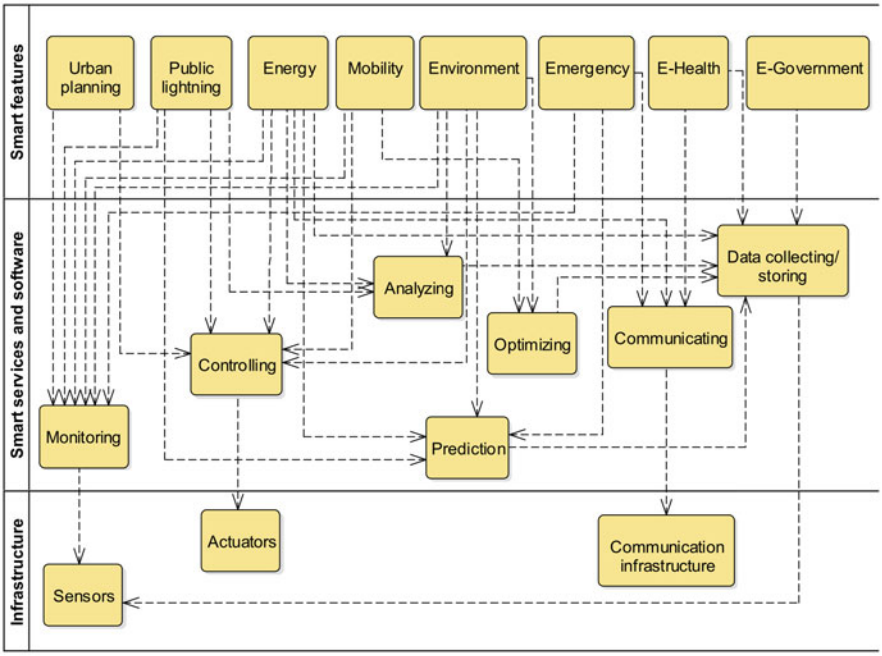


Fig. 10 Smart services across different application domains

elements of ICT services only into three levels. For better visualization, original levels 2 and 3 were merged together. At the lowest level, we decided to improve the readability by clustering the features and functionalities into the following hardware elements: Sensors, Actuators, and Communication infrastructure.

These elements are interconnected with middle-level elements. The middle level is represented with services and software providing the following functions: monitoring, controlling, analyzing, predicting, optimizing, communication, data collecting. More transparent overview of connections between top level and middle-level elements are depicted in Table 2. In Table 2, we distinguish between data collection and data storing. However, in Fig. 10, we consider it as one element on account of simplicity and because of the similarity of these two functions. Also, in Table 2, we consider that maintenance is not included in Fig. 10, since it is a separate function used only by one smart feature and without connections to the lowest level.

6 Reflections from Existing Smart Cities

In this section, we report our investigation on if and how the suggested methodology is applied in some of the current smart cities. In particular, we have explored the

Table 2 Smart services that are classified by software layers

Service and software layer	Smart services								Number occurence
	Urban planning	Energy	Mobility	Public lighting	Environment	Emergency	E-health	E-government	
Monitoring	*	*	*	*	*	*			6
Controling	*	*	*	*	*				5
Analysing		*		*	*				3
Predicting		*		*	*	*			4
Optimizing		*	*		*				3
Communication		*				*	*		3
Data storing		*					*	*	3
Maintenance		*							1
Data collecting		*							1

cities of Barcelona, Amsterdam, and New York, as well as some cities in Asia. In these cities, we have tried to recognize not only the features they are developing, but also the level of interconnection of the services and if the stakeholders in the city are supporting this kind of service cooperation. In [47], the reasons why Barcelona is evaluated on this position were summarized into three points:

- Use of ICT as a support of city processes for them to become more efficient, effective, transparent, and accessible;
- Interconnection of a city council, companies, educational institutions, and city inhabitants, which is called knowledge society;
- Processing, storing, and mainly sharing of information reflecting the fact that Barcelona provides Open Data from a public sector.

As an application of these three principles, we should mention the 22@ Barcelona project [138]. During detailed study of the project, we have found that the main emphasis was put on the massive application of ICT. It was possible to identify a positive effect for all city citizens, not only to those who were living in the area affected by the project, but also identifying added value the value, as it is co-created by people in the whole city. The description of the project is giving only brief information about combining different services together. Much of the project is presented as many separated services, sharing the same infrastructure and providing data that are classified mostly as open.

In Amsterdam, we have found general initiative, the so-called *Amsterdam smart city Challenge* [139]. This initiative has been collecting application proposals, based on the open data provided by the city. Also, they have identified opportunities driven by the data from applications and created a service environment. But most of the applications are designed as ad-hoc solutions and without any structured position.

In New York, the representatives of the city have already discovered the problem of service interconnection [140]. They are working on the new model of city development that could interconnect all the aspects of all stakeholders. An innovative approach is in usage of Minecraft-like software to model the future design of the street and model the behavior of the city components under different conditions. All the information, provided by this model, is taken as input to the discussion of the final design of the project.

In addition to the USA and Europe, we have found numerous innovative approaches in Asia, for instance, in Thailand. As one of the world's most famous touristic destinations, all cities of Thailand are motivated to increase their hospitality level also by applying smart city services. While in EU, and the USA, we have found the municipalities and public sector as one of the leaders of applying smart city services, we see different situation in Thailand. Its main activity is in hands of the private sector, mostly touristic agencies and big hotels. Therefore, the combination of the services used in Thailand is more flexible and focused on the main goals of the private leaders. Comparing to the EU, the value for the city citizens, from this perspective, is limited because the foreign tourists and guests are preferred. A typical example can be seen at www.phuket.com or www.bangkok.com. All pages are run by private companies, having very innovative design and offering modern services. On the other hand, official websites of municipalities (e.g., www.phuket.go.th and www.bangkok.go.th) are having very unpleasant design or are not working (Phuket website).

After investigating numerous examples of smart cities, we can identify the sharing of ICT infrastructure and promotion of open data. However, the attempts to provide added value via the service co-creation chain, from ICT services to the smart features, are rare and marginal, as a side effect of newly developed services. For instance, the investigated service was treated as possible sources of open data, without direct interaction with the services that might use their data.

7 Implications

The topics we deal with herein bring a number of intriguing implications for scholars and practitioners, particularly in terms of management and value co-creation process, there are relevant tips for the engagement (companies, municipalities—main providers) and self-engagement (people—main users) to be highlighted. From the managerial point of view, the reflections outlined here help us in understanding that there are different levels of connections in the pathways of smart city to be considered for the study of ICT services' provision effects, especially concerning peoples (citizens and others) directly involved as users.

The role of users in service design, during service provisions and after them (in terms of feedback) appear really active and functional for the positive level of performances inside the smart city, intended and analyzed as a whole [16, 125]. Users are the key resources that need to be integrated in the service provision [126], so with their involvement it could be possible to catch greater results than working alone [127, 128]. The will to include users is evident at all levels (and in many of presented projects), this is true in the information sharing, alert activities, database construction, access aspects, development of sensors, actuators and communication infrastructure and the improvement of all of phases of monitoring, controlling, analyzing, predicting, optimizing, communication, data collecting/storing.

Managers at all levels must be sensitive to this, by citizens and others to be part of the smart city in which they live in and operate [129], through special communication events, protocols diffusion, open days for demonstrations, differentiate newsletters, dedicate focus groups and so on. The lesson learned then could be in the sense of as much engagement as possible in order to include every interested actor in the provision process of ICT services in a smart city [130]. At the same time, the self-engagement of people is desirable, as they want to be more informed and actively participate to contribute effectively to improve the cited services in their smart city [130].

Value creation is an up-to-date topic for the worldwide service community nowadays. In the smart cities, following the exploited layer model, value is in the service usage [131, 132], and it is specifically identified in the context [133]. The advantages coming from their active and dynamically participation are especially in the greater value personally perceived from them (in terms of quality, speed, reliability, consistency, accuracy, punctuality). Value co-creation in smart cities [134] evolves the conditions of the local and global scenarios and leads scholars and practitioners to foster new proposals to increase co-creative processes, such as co-design, co-provision, co-marketing, co-delivery [135].

Nowadays, people are more habituated to the use of apps to control online advertisements, to suggest new kind of decisions for their daily life [136]. Thus, ICT services in a smart city derive at first from unavoidable cooperation between parties, more so if enabled by a diffused culture of value and right ready technologies [137]. This led to have many relationships, to foster them over time, also between users, through social, informal, or any else connections with each other. With the framework given by the layer model, a new frontier of value co-creation can be reached can be further investigated.

8 Conclusion

In this paper we have investigated the smart services that are developed in the domain of urban planning, energy, mobility, public lightning, environment, emergency, e-Health, and e-Government. For each domain we have organized the identified technologies and services into a smart city layered model, which contains the layers of infrastructure, software, supporting services, smart services, and smart features. Based on the results, we have discussed the findings in each domain, and derived the interpretations of technologies and services for each smart city layer across different domains. Furthermore, we have presented reflections on integrated approaches in existing smart cities, and discussed the implications from the managerial and value creation perspective. Overall, our analysis results can not only identify and understand the interconnections among technologies and services in the smart city, but also can be used to develop novel applications in smart cities from a holistic view. As future works, we plan to further investigate more smart city domains and interpret the findings of which domain-specific services can be used in other smart city domains as well as how to optimize the interconnections between technologies and service in each smart city layer.

Acknowledgments This research was supported by ERDF “CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence” (No. CZ.02.1.01/0.0/0.0/16_019/0000822).

References

1. Leonard Walletzký, Barbora Buhnova, and Luca Carrubbo. Value-driven conceptualization of services in the smart city: A layered approach. In *Social Dynamics in a Systems Perspective*, pages 85–98. Springer, 2018.

2. H. Bangui, B. Buhnova, S. Rakrak, and S. Raghay. Smart mobile technologies for the city of the future. In *smart city Symposium Prague*, pages 1–6. IEEE, 2017.
3. F. Caputo, V. Formisano, B. Buhnova, and L. Walletzký. Beyond the digital ecosystems view: Insights from smart communities. In *9th Annual Conference of the EuroMed Academy of Business*, 2016.
4. R.T. Golembiewski and G. Gabris. Tomorrow's city management: Guides for avoiding success-becoming-failure. *Public Administration Review*, page 240–246.
5. Milind Naphade, Guruduth Banavar, Colin Harrison, Jurij Paraszczak, and Robert Morris. Smarter Cities and Their Innovation Challenges. pages 32–39, 2011.
6. Yoshinari Shirai. Toward On-Demand Urban Air Quality Monitoring using Public Vehicles. pages 1–6, 2016.
7. California Public Utility Commission. Report to the governor and the legislature California, smart grid 2012. 2012.
8. D. Gesvindr, J. Michalkova, and B. Buhnova. System for collection and processing of smart home sensor data. In *Software Architecture Workshops (ICSAW), 2017 IEEE International Conference on*, pages 247–250. IEEE, 2017.
9. Sadia Sharmin, and Sikder Tahsin Al-Amin. A cloud-based dynamic waste management system for smart cities. In Proceedings of the 7th Annual Symposium on Computing for Development, pp. 1–4. 2016.
10. Smart Grid Standardization. The German Standardization Roadmap E-Energy/Smart Grid, 2010.
11. Wayne Eckerson. Smart companies in the 21st century: The secrets of creating successful business intelligence solutions. *TDWI Report Series*, 7, 2003.
12. Luis Montestruque, South Bend, Fitzpatrick Hall, and Notre Dame. Globally coordinated distributed storm water management system. In Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks, pp. 1–6. 2015.
13. Cengiz Gezer and Chiara Buratti. A ZigBee Smart Energy Implementation for Energy Efficient Buildings. 2011.
14. Shirin Noei and Arman Sargolzaei. Reducing Traffic Congestion Using Geo-fence Technology: Application for emergency Car. pages 15–20, 2014.
15. Zubaida Alazawi, Omar Alani, and Mohmmad B Abdljabar. A Smart Disaster Management System for Future Cities. pages 1–10, 2014.
16. R.C. Basole and W.B Rouse. Complexity of service value networks: Conceptualization and empirical investigation. In *IBM Systems Journal*, volume 47, page 1–53.
17. Cen-cenelec-etsi smart grid and Coordination Group. CEN-CENELEC-ETSI smart grid Coordination Group smart grid Reference Architecture Contents. (November), 2012.
18. Jörn Trefke, Sebastian Rohjans, Mathias Uslar, Sebastian Lehnhoff, L Nordstrom, and Arshad Saleem. smart grid architecture model use case management in a large European smart grid project. In *Innovative smart grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*, pages 1–5. IEEE, 2013.
19. Azedine Charef, Antoine Ghauch, Patrick Baussand, and Michel Martin-bouyer. Water quality monitoring using a smart sensing system. 28:219–224, 2000.
20. J.A. Coca-Stefaniak, C. Parker, S. Quin, R. Rinaldi, and J. Byrom. Town centre management models: A European perspective. *Cities*, 26(2):74–80.
21. Terézia Kazičková and Barbora Bühnová. Ict architecture for the smart grid: Concept overview. In *smart cities Symposium Prague, 2016*, pages 1–4. IEEE, 2016.
22. Roli Shukla. Smart waste management. 5(2), 2016.
23. Leonidas Anthopoulos and Marijn Janssen. Smart Service Portfolios: Do the Cities Follow Standards pages 357–362, 1895.
24. G Piro, I Cianci, L A Grieco, G Boggia, and P Camarda. The Journal of Systems and Software Information centric services in smart cities. *The Journal of Systems & Software*, 88:169–188, 2014.
25. Sherin Abdelhamid and Hossam S Hassanein. On the Provisioning of Vehicle-Based Public Sensing Services. pages 71–77.

26. Istituto Superiore, Mario Boella, Claudio Rossi, Manuel Gaetani, and Antonio Defina. AURORA: an Energy Efficient Public Lighting IoT System for Smart Cities. 44(2):76–81, 2016.
27. F. Caputo, B. Buhnova, and L. Walletzký. Investigating the role of smartness for sustainability: insights from the smart grid domain. *Sustainability Science*, pages 1–11, 2018.
28. Matthias D Galus, Rashid A Waraich, Fabrizio Noembrini, Karel Steurs, Gil Georges, Konstantinos Boulouchos, Kay W Axhausen, and Göran Andersson. Integrating Power Systems , Transport Systems and Vehicle Technology for Electric mobility Impact Assessment and Efficient Control. 3(2):934–949, 2012.
29. Marie Anne Macadar, Jorge Lheureux-de-Freitas, Luiza Schuch de Azambuja, and Edimara Mezzomo Luciano. Contact Center in a smart cities View: a Comparative Case Study of Curitiba (Brazil), Porto Alegre (Brazil) and Philadelphia (USA). 2016, pp. 215–222.
30. Gang Pan, Guande Qi, Wangsheng Zhang, Shijian Li, and Zhaohui Wu. Trace Analysis and Mining for smart cities: Issues , Methods , and Applications. (June):120–126, 2013.
31. Irene Celino and Cefriel Politecnico. Smart Urban Planning Support through Web Data Science on Open and Enterprise Data. pages 1407–1412.
32. Xuxu Chen, Yu Zheng, Yubiao Chen, Qiwei Jin, Weiwei Sun, Eric Chang, and Wei-ying Ma. Indoor Air Quality Monitoring System for Smart Buildings. pages 2–6, 2014.
33. Nikolaos Rapousis, Michalis Katsarakis, and Maria Papadopouli. QoWater - A crowd-sourcing approach for assessing the water quality.
34. Catriona Manville, Gavin Cochrane, Jonathan Cave, Jeremy Millard, Jeremy Kevin Pederson, Rasmus Kåre Thaarup, Andrea Liebe, Matthias Wissner, Roel Massink, and Bas Kotterink. Mapping smart cities in the EU. 2014.
35. Authors Catalin, Felix Covrig, Mircea Ardelean, Julija Vasiljevska, Manuel Sánchez Jiménez, Constantina Filiou, and D G Ener. *smart grid Projects Outlook 2014*. 2014.
36. H. Farhangi. The Path of the smart grid. *In IEEE Power & Energy Magazine*, 8(February), 2010.
37. Vincenzo Giordano, Flavia Gangale, Gianluca Fulli, M. Sánchez Jiménez, I. Onyeji, A. Colta, I. Papaioannou et al. Smart grid projects in Europe: lessons learned and current developments, JRC Reference Reports, Publications Office of the European Union, 2011.
38. Isaac Ghansah. smart grid cyber security potential threats, vulnerabilities and risks. *California Energy Commission, PIER Energy-Related environmental Research Program, CEC-500-2012-047*, 2009.
39. Punya Prakash. Data concentrators: The core of energy and. 2012
40. Robert Johnson. Smart Metering: A review of Smart Metering and Survey options for Energy. 2010, no. September.
41. R. Hambleton. The new city management. In *Globalism and Local Democracy*, page 147–168. Palgrave Macmillan, London.
42. Agnes Mainka, Sarah Hartmann, and Wolfgang G Stock. Open Innovation in Smart Cities: Civic Participation and Co-Creation of Public Services. 2015.
43. J. Tsado, O. Imoru, and O. D. Segun. System Stability Enhancement through smart grid Technologies with DRS. 2012, vol. 2, no. 4, pp. 621–629.
44. Freescale. Smart Energy. Industrial brochure: www.nxp.com/assets/ documents/data/en/brochures/BRSMRTENERGY.pdf. 2013.
45. Jixuan Zheng, Li Lin. Smart Meters in smart grid: An Overview. 2013. ISBN 9780769549668. Available from DOI: 10.1109/GreenTech. 2013.17.
46. Flavia Gangale, Gianluca Fulli, and Manuel Sánchez Jiménez Vincenzo Giordano. Smart Grid projects in Europe: lessons learned and current developments. European Commission Joint Research Centre Institute for Energy, Luxembourg, Reference Report ISBN (2013): 978-92.
47. Petr Štěpánek, Mouzhi Ge, and Leonard Walletzky. It-enabled digital service design principles - lessons learned from digital cities. In Marinos Themistocleous and Vincenzo Morabito, editors, *Information Systems*, pages 186–196, Cham, 2017. Springer International Publishing.

48. Martin Böhm, Barbara Flechl, Alexander Frötscher, Alexander Hausmann, Wolfgang Kernstock, Claus Seibt, Lucas Weiss, and Katharina Zwick. *ICT concepts for optimization of mobility in smart cities FINAL REPORT A study prepared for the European Commission*.
49. So Hashimoto, Ryo Kanamori, Takayuki Ito, and Shantanu Chakraborty. Evaluation of parking reservation system with auction including electricity trading. pages 389–393, 2012.
50. Sara Deilami, Amir S. Masoum, Paul S. Moses, and Mohammad AS Masoum. Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile. IEEE Transactions on Smart Grid 2, no. 3 (2011): 456–467.
51. David Miller. Does The First Officer Concur ? Shared Control with Smart Vehicle Systems Abstract. 2014. ISBN 978-1-45030-725-3.
52. Sukru Yaren Gelbal, Bilin Aksun Guvenc, and Levent Guvenc. SmartShuttle: A unified, scalable and replicable approach to connected and automated driving in a smart city. In Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering, pp. 57–62. 2017.
53. Margarita Angelidou. Smart city policies: A spatial approach. Cities 41 (2014): S3-S11.
54. Junghoon Lee, Gyung-Leen Park, Hye-Jin Kim, and Hongbeom Jeon. Fast scheduling policy for electric vehicle charging stations in smart transportation. In Proceedings of the 2011 ACM Symposium on Research in Applied Computation, pp. 110–112. 2011.
55. Päivi Eriksson, and Hanna Lehtimäki. Strategy rhetoric in city management: How the presumptions of classic strategic management live on?. Scandinavian Journal of Management 17, no. 2 (2001): 201–223.
56. Istituto Superiore, Mario Boella, Claudio Rossi, Manuel Gaetani, Antonio Defina. AURORA: an Energy Efficient Public Lighting IoT System for smart cities. 2016, vol. 44, no. 2, pp. 76–81.
57. Roli Shukla. Smart waste management. 2016, vol. 5, no. 2
58. Ingo von Poser, and Adel R. Awad. Optimal routing for solid waste collection in cities by using real genetic algorithm. In 2006 2nd International Conference on Information & Communication Technologies, vol. 1, pp. 221–226. IEEE, 2006.
59. Sadia Sharmin. A Cloud-based Dynamic Waste Management System for smart cities, 2000. ISBN 978-1-45034-649-8.
60. Yoshinari Shirai. Toward On-Demand Urban Air Quality Monitoring using Public Vehicles. 2016, pp. 1–6. ISBN 978-1-45034-667-2.
61. Nikolaos Rapousis, Michalis Katsarakis, and Maria Papadopouli. QoWater: A crowd-sourcing approach for assessing the water quality. In Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks, pp. 1–6. 2015.
62. Azedine Charef, Antoine Ghauch, Patrick Baussand, and Michel Martin-Bouyer. Water quality monitoring using a smart sensing system. Measurement 28, no. 3 (2000): 219–224.
63. Luis Montestruque, and M. D. Lemmon. Globally coordinated distributed storm water management system. In Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks, pp. 1–6. 2015.
64. Shirin Noei, Hugo Santana, Arman Sargolzaei, and Mohammadreza Noei. Reducing traffic congestion using geo-fence technology: application for emergency car. In Proceedings of the 1st International Workshop on Emerging Multimedia Applications and Services for Smart Cities, pp. 15–20. 2014.
65. Zubaida Alazawi, Omar Alani, Mohmmad B. Abdljabar, Saleh Altowaijri, and Rashid Mehmood. A smart disaster management system for future cities. In Proceedings of the 2014 ACM international workshop on wireless and mobile technologies for smart cities, pp. 1–10. 2014.
66. Anthony Maeder. e-Research Meets e-Health. 80(Hdkm):3–6, 2008.
67. Henna Kim and Bo Xie. Health Literacy and Internet- and Mobile App-based Health Services:A Systematic Review of the Literature. 2015.
68. Song Han, Geoff Skinner, Vidyasagar Potdar, and Elizabeth Chang. A Framework of Authentication and Authorization for e-Health Services. pages 105–106, 2006.

69. Xingguang Zhou, Jianwei Liu, Weiran Liu, and Qianhong Wu. Anonymous role-based access control on e-health records. In Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security, pp. 559–570. 2016.
70. Dan Johansson and Mikael Wiberg. Mobile e-Services and Open Data in e-Government Processes - Transforming Citizen Involvement Categories and Subject Descriptors. 2015.
71. Sehl Mellouli, Luis F Luna-reyes, and Jing Zhang. Smart government , citizen participation and open data. 19:1–4, 2014.
72. Leonidas Anthopoulos and Christopher G Reddick. smart city and Smart Government: Synonymous or Complementary? pages 351–355, 2016.
73. Bernadia Irawati Tjandradewi, and Peter J. Marcotullio. City-to-city networks: Asian perspectives on key elements and areas for success. Habitat International 33, no. 2 (2009): 165–172.
74. L.J.G. Gutiérrez, I.T. Torres, and V.B. Molina. Quality management initiatives in Europe: An empirical analysis according to their structural elements. *Total Quality Management*, 21(6):577–601, 2010.
75. H. George Frederickson. The Adapted City: Institutional Dynamics and Structural Change: Institutional Dynamics and Structural Change. Routledge, 2016.
76. M. Gandy. Rethinking urban metabolism: water, space and the modern city. *City*, 8(3):363–379, 2004.
77. J.R. Kenworthy. The eco-city: ten key transport and planning dimensions for sustainable city development. *environment and*, 18(1):67–85, 2006.
78. H. Quak and L. Tavasszy. Customized solutions for sustainable city logistics: the viability of urban freight consolidation centres. *In Transitions towards sustainable mobility*, page 213–233, 2011.
79. R.J. Rogerson. Quality of life and city competitiveness. *Urban studies*, 36(5–6):969–985, 1999.
80. D. Driskell. *Creating better cities with children and youth: A manual for participation*. Routledge, 2017.
81. C. Landry. The creative city: A toolkit for urban innovators, 2012.
82. M. Bruce, M. Peltu, and W.H. Dutton. *Society on the line: Information politics in the digital age*. Oxford University Press, 1999.
83. J. Musso, C. Weare, and M. Hale. Designing web technologies for local governance reform: Good management or good democracy? *Political Communication*, 17(1):1–19, 2000.
84. A. Aurigi. New technologies, same dilemmas: policy and design issues for the augmented city. *Journal of urban technology*, 13(3):5–28, 2006.
85. M. Gurstein, editor. *Community Informatics: Enabling Communities with Information and Communications Technologies: Enabling Communities with Information and Communications Technologies*. IGI Global, London, 1999.
86. M.J. Moon. The evolution of e-government among municipalities: rhetoric or reality? *Public administration review*, 62(4):424–433, 2002.
87. N. Odendaal. Information and communication technology and local governance: understanding the difference between cities in developed and emerging economies. *Computers, environment and Urban Systems*, 27(6):585–607, 2003.
88. B.N. Hague and B. Loader, editors. *Digital democracy: Discourse and decision making in the information age*. Psychology Press, London, 1999.
89. Caputo F., Walletzky L., and Mouzhi G. Modelling the service value chain for smart city. In *IDIMT 2017: Interdisciplinary Information Management Talks, Podebrady, Czech Republic, September*, page 6–8, 2017.
90. Carayannis E.G., Caputo F., and Del Giudice M. Technology transfer as driver of smart growth: A quadruple/quintuple innovation framework approach. in vrontis. page 295–315. EuroMed Pres, Cyprus, 2017.
91. Carayannis G.E., Del Giudice M., Saviano M., and Caputo F. Beyond big data: From smart to wise knowledge management. In Barile S., Espejo R., Perko I., and Saviano M, editors, *Cybernetics and Systems. Social and Business Decisions*. Giappichelli-Routledge, London, 2018.

92. M. Batty. Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3):274–279, 2013.
93. A. Cocchia. Smart and digital city: A systematic literature review. In *smart city*, page 13–43. Springer, Cham, 2014.
94. A. Macintosh. Characterizing e-participation in policy-making. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, page 10. IEEE, 2004.
95. H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira. smart cities and the future internet: Towards cooperation frameworks for open innovation, 2011. In *The future internet assembly*, page 431–446. Springer, Berlin, Heidelberg.
96. F. Caputo and L. Walletzky. Investigating the users' approach to ict platforms in the city management. *Systems*, 5(1):1, 2017.
97. D. Gichoya. Factors affecting the successful implementation of ict projects in government. *the Electronic Journal of e-government*, 3(4):175–184, 2005.
98. M. Castro, A.J. Jara, and A.F. Skarmeta. Smart lighting solutions for smart cities. In *Advanced information networking and applications workshops (WAINA), 2013 27th international conference on*, page 1374–1379. IEEE, 2013.
99. F. Bifulco, M. Tregua, C.C. Amitrano, and A. D'Auria. Ict and sustainability in smart cities management. *International Journal of Public Sector Management*, 29(2):132–147, 2016.
100. R.G. Hollands. Will the real smart city please stand up? intelligent, progressive or entrepreneurial? *City*, 12(3):303–320, 2008.
101. T. Bakici, E. Almirall, and J. Wareham. A smart city initiative: the case of Barcelona. *Journal of the Knowledge Economy*, 4(2):135–148, 2013.
102. K.A. Paskaleva. Enabling the smart city: The progress of city e-governance in Europe. *International Journal of Innovation and Regional Development*, 1(4):405–422, 2009.
103. J.M. Hernández-Munoz, J.B. Vercher, L. Munoz, J.A. Galache, M. Presser, L.A.H. Gómez, and J. Pettersson. smart cities at the forefront of the future internet. In *The future internet assembly*, page 447–462. Springer, Berlin, Heidelberg, 2011.
104. H. Chourabi, T. Nam, S. Walker, J.R. Gil-Garcia, S. Mellouli, K. Nahon, and H.J. Scholl. Understanding smart cities: An integrative framework. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, page 2289–2297. IEEE, 2012.
105. The role of advanced sensing in smart cities. *Sensors*, 13(1):393–425, 2012.
106. P. Lombardi, S. Giordano, H. Farouh, and W. Yousef. Modelling the smart city performance. *Innovation: The European Journal of Social Science Research*, 25(2):137–149, 2012.
107. smart cities: governing, modelling and analysing the transition, 2013.
108. C.E. Mulligan and M. Olsson. Architectural implications of smart city business models: An evolutionary perspective. *IEEE Communications Magazine*, 51(6):80–85, 2013.
109. T. Nam and T.A. Pardo. 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*, page 185–194, 2011.
110. L.G. Anthopoulos and A. Vakali. Urban planning and smart cities: Interrelations and reciprocities. In *The Future Internet Assembly*, page 178–189. Springer, Berlin, Heidelberg, 2012.
111. J.H. Lee, R. Phaal, and S.H. Lee. An integrated service-device-technology roadmap for smart city development. *Technological Forecasting and Social Change*, 80(2):286–306, 2013.
112. L. Anthopoulos and P. Fitsilis. Exploring architectural and organizational features in smart cities. In *Advanced Communication Technology (ICACT), 2014 16th International Conference on*, page 190–195. IEEE, 2014.
113. L. Leydesdorff and M. Deakin. The triple-helix model of smart cities: A neo-evolutionary perspective. *Journal of urban technology*, 18(2):53–63, 2011.
114. A. Vanolo. Smartmentality: The smart city as disciplinary strategy. *Urban Studies*, 51(5):883–898, 2014.
115. Saviano M., Caputo F., Formisano V., and Walletzky L. From theory to practice: applying systems thinking to smart cities. *Business Systems. E-book*, page 35–40, 2016.

116. Saviano M., Barile S., and Caputo F. Re-affirming the need for systems thinking in social sciences: A viable systems view of smart city. in vrontis. page 1552–1567. EuroMed Pres, Cyprus, 2017.
117. Caputo F., Buhnova B., and Walletzky L. Investigating the role of smartness for sustainability: Insights from the smart grid domain. *Sustainability Science*, 2018.
118. Caputo F., Buhnova B., and Walletzky L. Reflecting upon the role of smart grid in linking smartness and sustainability. In Barile S., Espejo R., Perko I., and Saviano M, editors, *Cybernetics and Systems. Social and Business Decisions*. Giappichelli-Routledge, London, 2018.
119. L. Anthopoulos, M. Janssen, and V. Weerakkody. Smart service portfolios: Do the cities follow standards? In *Proceedings of the 25th International Conference Companion on World Wide Web*, page 357–362. International World Wide Web Conferences Steering Committee, 2016.
120. S. Musa. smart cities-a road map for development. *IEEE Potentials*, 37(2):19–23, 2018.
121. Caputo F.Walletzky L. and Stepanek P. How smart technologies and big data affect systems' lives? conceptual reflections on the smart city'. In ecosystem. In Barile S., Espejo R., Perko I., and Saviano M, editors, *Cybernetics and Systems. Social and Business Decisions*. Giappichelli-Routledge, London, 2018.
122. A.Y. Saber and G.K. Venayagamoorthy. Plug-in vehicles and renewable energy sources for cost and emission reductions. *IEEE Transactions on Industrial electronics*, 58(4):1229–1238, 2011.
123. M.D. Galus, R.A. Waraich, F. Noembrini, K. Steurs, G. Georges, K. Boulouchos, and G. Andersson. Integrating power systems, transport systems and vehicle technology for electric mobility impact assessment and efficient control. *IEEE Transactions on smart grid*, 3(2):934–949, 2012.
124. M.A. Macadar, J. Lheureux-de Freitas, L.S. de Azambuja, and E.M. Luciano. Contact center in a smart cities view: a comparative case study of Curitiba (Brazil). In *Porto Alegre (Brazil) and Philadelphia (USA). In Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance*, page 215–222, 2016.
125. Mele W.B. and F. Polese. Key dimensions of service systems: Interaction in social & technological networks to foster value co-creation. In V. Krishna (A H. Demirkan, J. Spohrer, editor, *CURA DI), The Science of Service Systems*. Springer, pp.37–59, Springer, New York, 2011.
126. H. Wieland, F. Polese, S. Vargo, and R. Lusch. Toward a service (eco) systems perspective on value creation. int. *J. Serv. Sci. Manag. Eng. Technol*, 2012(3):12–24, 2012.
127. Prahalad C.K. and Ramanswamy V. *The future of competition: Co-creating unique value with customers*. Harvard University Press, Cambridge, MA, 2004.
128. S.L. Vargo, P.P. Maglio, and M.A. Akaka. On value and value co-creation: A service systems and service logic perspective. *Eur. Manag. J*, 2008:26–145, 2008.
129. L. Carrubbo, D. Sarno, F. Caputo, and F. Smaldone. "managing value dimensions within an ecosystem framework: reflections and empirical observations in the tourism sector" in African. *Journal of Business Management*, 12:24–713, 2018.
130. F. Polese and Carrubbo L. Eco-sistemi di servizio in sanità, collana studi e. 2016.
131. Ballantyne D. and R.J Varey. Creating value-in-use through marketing interaction: the exchange logic of relating, communicating and knowing. In *Marketing Theory n.3*, page 335–348, 2006.
132. S.L. Vargo and R.F. Lusch. From repeat patronage to value co-creation in service ecosystems: A transcending conceptualization of relationship. *J. Bus. Mark. Manag*, 2010(4):169–179, 2010.
133. Vargo S.L. Chandler, J.D. Contextualization and value-in-context: How context frames exchange. In *Marketing Theory*, volume 11, page 1–35, 2011.
134. S. Barile, F. Polese, L. Carrubbo, F. Caputo, and Wallerzky L. "determinants for value co-creation and collaborative paths in complex service systems: A. In *Focus on (Smart) Cities", in Service Science*, volume 10, page 4–379, 2018.

135. O. Troisi, Carrubbo L, G. Maione, and C. Torre. "unity is strength: Co-creating value in working context", in. *Journal of Business and Economics*, 8:1–97, 2017.
136. Polese F., Troisi O., L. Carrubbo, Grimaldi M., and Monda A. *Technology in value co-creation experiences: how ICTs shape customer*. McGraw-Hill Education, 2018.
137. F. Iandolo, S. Barile, S. Armenia, and L. Carrubbo. System dynamics perspective on a viable systems approach definition for sustainable value. In *Sustainability Science*. Ed. Springer, 2018.
138. ECPA Urban Planning. Case Study: 22@ Barcelona Innovation District, 2017. URL https://www.smartcitiesdive.com/ex/sustainablecitiescollective/case-study-22-barcelona-innovation-district/27601/
139. AMS Institute. Amsterdam City Challenge: Urban Mobility & Sustainable Energy, 2019. URL https://amsterdamsmartcity.com/updates/event/amsterdam-sustainable-city-challenge
140. Smart City. The Equitable City – A New Name For New York, 2017. URL https://smartcity.press/new-yorks-smart-city-initiatives/

A Novel Data Representation Method for Smart Cities' Big Data

Attila M. Nagy and Vilmos Simon

1 Introduction

The rapid development of smart city information and communication technologies provided many new public city services to the city dwellers, also improving the effectiveness of existing services.

The appearance of Intelligent Transportation Systems (ITSs) provided the opportunity to integrate traffic data collection and analysis with transportation intervention methods (through traffic lights, Vehicle-to-Everything (V2X) systems, city apps), constituting a fully functioning, real-time, and efficient transportation management framework [1]. For example, a traffic signal system [2] can modify the signal phase settings adaptively to the current traffic state, which was measured with the help of deployed sensor networks or cameras. From these devices, traffic time series data can be extracted, which are necessary to find the right traffic light settings [3]. Smart parking solutions [4] should be part of ITSs as a remarkable fraction of the overall motorist travel time is often spent on the search of a free parking lot in the city. This search time can be reduced by using the historical parking lot occupancy time series from the parking spots in the city, providing an accurate prediction of the number of available parking spots near at hand. These predictions help to reduce congestion near the parking sites, but collection and processing of huge time series datasets are required. The effectiveness of aforementioned systems can be measured in several ways; one is the continuous monitoring of pollution levels [5] in urban areas, where time series data extracted from air pollution sensors are used to measure the emission of the pollutants. With the fast evaluation of these time

A. M. Nagy (✉) · V. Simon
Budapest University of Technology and Economics, Faculty of Electrical Engineering and Informatics, Department of Networked Systems and Services, Budapest, Hungary
e-mail: anagy@hit.bme.hu; svilmos@hit.bme.hu

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_6

series, basic information is gathered to take further actions to reduce the pollution level.

Parallel to cities' infrastructure, devices of the city dwellers are developing as well. An exciting example is Human Activity Recognition (HAR). HAR uses smartphones equipped with accelerometer and gyroscope to provide rich three-dimensional motion data [6]. The time series data from these sensors can be used for various purposes such as Smart Healthcare [7], daily activity classification [8], or elder care [9].

As the previous examples showed, time series datasets are generated from many different data sources. However, data scientists face the difficulty of efficient processing of these tremendous time series datasets [10], as they are used in various smart city-related data mining fields like clustering [11], classification [12], or anomaly detection [13]. These datasets generally have high dimensionality and feature correlation. Furthermore, they also contain a large amount of noise, which typically comes from inaccurate or erroneous measurements.

The high dimensionality of such time series increases both the access time to the data and the computation time needed for classification models or making clusters. Furthermore, visualization techniques need to employ data reduction and aggregation techniques to cope with the high volume of data that cannot be plotted in detail at once [14]. In order to speed up data mining tasks and reduce their storage space demand, many high-level representations of the raw time series data have been proposed utilizing dimensionality or numerosity reduction.

The well-known representations are Discrete Fourier Transform (DFT) [15], Discrete Wavelet Transform (DWT) [16], PAA [17, 18], Adaptive Piecewise Constant Approximation (APCA) [19], Singular Value Decomposition (SVD) [20], or SAX [21]. All of these representations have their own strengths and weaknesses, and therefore choosing the appropriate representation always depends on the dataset and the scope of the work.

When the main goal is the identification of particular behaviors, pattern recognition, or anomaly detection, one of the most prominent high-level representation techniques is the SAX [21] representation in which the subsequences of the whole time series are represented by symbols. Besides that, it also provides a distance measure that makes possible to calculate the distance between two distinct sequences of symbols. Many extensions of SAX are available in the literature [22, 14, 23, 24, 25], but none of them focuses on the time series which have multiple variables.

Multivariate time series data is ubiquitous and broadly available in smart cities, since in many applications the investigated process is described by more than one variable. For instance, a traffic monitoring sensor typically measures the traffic flow, occupancy, and speed variables simultaneously [26], or smartphones equipped with accelerometer and gyroscope provide rich three-dimensional motion data [6]. If the original SAX method is applied on the distinct variables, a new symbol sequence for each variable is provided. The problem with this approach is that in many cases only the joint examination of these variables could provide appropriate results, because one variable does not contain enough information about the investigated process.

In this chapter, a multivariate extension of SAX will be presented which allows to represent multivariate time series with one sequence of symbols. The resulting symbol sequence has the same equiprobability property as the original SAX representation and it also guarantees the same minimum distance measure.

Another challenge of using SAX is the appropriate setting of its parameters [27]. Assuming that the available dataset is composed of M time series of length N, the empirical parameter setting presented in [21] executes $\mathcal{O}(M^2)$ calculations for each parameter setting. The procedure proposed in [21] is a computationally intense task, because it tries with all possible parameter settings for all the $M^2 - M$ time series pairs. Unfortunately, the SAX does not allow the usage of well-known gradient descent optimization procedure to find the best parameters, as the error function is not continuous (caused by the symbolic transformation). In this chapter, we propose a new parameter optimization method for SAX, which uses derivative-free optimization algorithms to find the best parameter setting.

The rest of this chapter is organized as follows. In Sect. 2, the original SAX, its extension, and SAX parameter optimization techniques will be presented. Our multivariate method is introduced in Sect. 3. In Sect. 4, an improved distance method is presented which is suitable for multivariate environment. Our parameter optimization method is proposed in Sect. 5, while in Sect. 6, our algorithm is applied on several datasets to measure its performance. Section 7 briefly summarizes our results.

2 Background and Related Works

The SAX offers dimensionality and numerosity reduction, and it provides a distance measure between symbol sequences that bounds from below the Euclidean distance on the original time series. The lower bounding distance measures, which underestimate the Euclidean distance, are highly used in huge time series databases. Using these cheap SAX estimations, a set of candidate time series can be found, which is extremely useful for fast and efficient searching in the database.

2.1 Time Series Normalization

The examined time series is normalized in the first step of the SAX. This step is necessary, because without normalization the result is meaningless [28].

Let $\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$ denote a time series with length N. To normalize $\boldsymbol{X}$, the experimental mean, μ, and the experimental standard deviation, σ, have to be calculated first:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} X_n, \tag{1}$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \mu)^2}. \tag{2}$$

Based on μ and σ, the normalized time series, $\hat{\boldsymbol{X}}$, is obtained by

$$\hat{X}_n = \frac{X_n - \mu}{\sigma}, \quad n = 1, 2, \ldots, N. \tag{3}$$

The experimental mean and the experimental standard deviation of $\hat{\boldsymbol{X}}$ are 0 and 1, respectively.

2.2 Dimensionality Reduction via PAA

In the next step of SAX, the PAA dimensionality reduction technique is applied on the normalized $\hat{\boldsymbol{X}}$ time series. The advantage of the PAA method is its simplicity and clear intuitive meaning, while it has been shown to rival more sophisticated dimensionality reduction techniques like DFT and DWT [29].

The PAA transforms an $\hat{\boldsymbol{X}}$ normalized time series to an $\check{\boldsymbol{X}} = \{\check{X}_1, \check{X}_2, \ldots, \check{X}_K\}$ time series with length K ($K << N$). The original $\hat{\boldsymbol{X}}$ time series is divided into K partitions. In each partition, the sum of the values is divided by their mean. The kth element of the resulting $\check{\boldsymbol{X}}$ time series can be obtained as

$$\check{X}_k = \frac{1}{\left\lceil \frac{N}{K} \right\rceil} \sum_{j=\left\lceil \frac{N}{K} \right\rceil (k-1)+1}^{\left\lceil \frac{N}{K} \right\rceil k} \hat{X}_j, \quad k = 1, 2, \ldots, K-1, \tag{4}$$

$$\check{X}_K = \frac{1}{N - \left\lceil \frac{N}{K} \right\rceil (K-1)} \sum_{j=\left\lceil \frac{N}{K} \right\rceil (K-1)+1}^{N} \hat{X}_j.$$

If N is an integer multiple of K, then (4) simplifies to

$$\check{X}_k = \frac{K}{N} \sum_{j=\frac{N(k-1)}{K}+1}^{\frac{Nk}{K}} \hat{X}_j, \quad k = 1, 2, \ldots, K. \tag{5}$$

2.3 Discretization Step

The discretization step is the last step of the SAX method, which results in a symbolic sequence. Since the normalized time series have the standard normal distribution [30], the equiprobability [31] property can be easily ensured by defining appropriate breakpoints.

Let $\mathcal{A} = \{A_1, A_2, \ldots, A_{|\mathcal{A}|}\}$ be an alphabet of size $|\mathcal{A}|$ and $F(x)$ the cumulative distribution function of the standard normal distribution. The equiprobability breakpoints are $B = \{\beta_1, \beta_2, \ldots, \beta_{|\mathcal{A}|-1}\}$, where

$$F(\beta_a) = \frac{a}{|\mathcal{A}|}, \quad a = 1, \ldots, |\mathcal{A}| - 1.$$

By using these breakpoints, a sample $\check{X}_k$ from the $\check{\boldsymbol{X}}$ time series, with $\beta_a \leq \check{X}_k < \beta_{a+1}$, is mapped to symbol A_a. That is, the discretization step results in the $\tilde{\boldsymbol{X}} = \{\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_K\}$ time series, where $\tilde{X}_k$ is the A_a symbol of the alphabet when $\beta_a \leq \check{X}_k < \beta_{a+1}$.

2.4 Distance Measure

In [21], the symbolic representation of the SAX method is supplied with the $MINDIST$ distance measure, defined as

$$MINDIST(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}) = \sqrt{\frac{1}{K}\sum_{k=1}^{K} dist(\tilde{X}_k, \tilde{Y}_k)^2}, \tag{6}$$

where $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{Y}}$ are symbol sequences of the same alphabet $\mathcal{A}$ of size K, and the $dist(A_r, A_c)$ distance between symbols A_r and A_c of alphabet $\mathcal{A}$ is defined as

$$dist(A_r, A_c) = \begin{cases} 0 & \text{if } |r - c| \leq 1, \\ \beta_{c-1} - \beta_r & \text{if } r < c - 1, \\ \beta_{r-1} - \beta_c & \text{if } c < r - 1. \end{cases} \tag{7}$$

The authors of [21] also proved that the Euclidean distance, defined by

$$D_{Euclidean}(\boldsymbol{X}, \boldsymbol{Y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_n - Y_n)^2}, \tag{8}$$

is an upper bound of $MINDIST$ for all K and $|\mathcal{A}|$ parameter combinations.

2.5 Parameter Setting

The SAX method has two parameters to be set, K and $|\mathcal{A}|$. K denotes the number of partitions per time series and $|\mathcal{A}|$ is the size of alphabet $\mathcal{A}$. When approximating a massive dataset with SAX, these parameters have to be chosen in such a way that the symbolic representation makes the best use of the available primary memory [21]. To find the best parameter settings, the authors of [21] empirically evaluated the tightness of the lower bound, defined by

$$\text{TLB}(K, |\mathcal{A}|) = \frac{MINDIST(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}})}{D_{Euclidean}(\boldsymbol{X}, \boldsymbol{Y})}, \tag{9}$$

where $\tilde{\boldsymbol{X}}$ ($\tilde{\boldsymbol{Y}}$) is obtained from $\boldsymbol{X}$ ($\boldsymbol{Y}$) using K partitions and an alphabet of $|\mathcal{A}|$ symbols. The memory requirement of a parameter setting is characterized by $K \log(|\mathcal{A}|)$, and the best parameter set for a given available memory mem is

$$\underset{K, |\mathcal{A}| : K \log(|\mathcal{A}|) \leq mem}{\arg\min} \text{TLB}(K, |\mathcal{A}|). \tag{10}$$

2.6 Extensions of the SAX Method

The SAX representation has its own limitation. Since the SAX uses the mean values of partitions for symbol mapping, it does not take into account the trends (or directions) inside the partition. Because of this, it happens sometimes that $MINDIST$ between two partitions is zero based on the mean values, while the two partitions have completely different trends. To ease this limitation, extensions of the SAX method are proposed to eliminate the distortion by the mean values. Table 1 shows some examples for the modified SAX symbols.

A trend-based approximation, indicating the trend for each partition, is proposed in [14]. The three considered trend types are up (U), down (D), and straight (S). The trend type is determined by least square fitting of a line on the data points within a partition. The slope of the best fitting line characterizes the trend of a partition which is attached to the symbol of the partition. If the absolute value of the steepness of the slope is smaller than an ϵ threshold, it is considered as straight (S), otherwise it can be up (U) or down (D) based on the positive or negative sign of the steepness. The disadvantage of the method is that it is not suitable for distance calculation, and however it can be used for classification purposes.

Similarly, the Trend based Symbolic Aggregate Approximation (TSAX) extends the SAX method with trend information [32]. TSAX adds two so-called trend indicators to each partition. To determine trend indicators, the partition is divided into three parts. Then, the relation of mean values of the consecutive parts is determined, which results in two trend indicators. The first trend indicator describes

Table 1 Extensions of the SAX method

Method	Symbol series	Ref.
SAX	$a\ a\ d\ c\ b\ a$	[21]
Trend-based SAX_1	$a_U\ a_U\ d_D\ c_S\ b_S$	[14]
Trend-based SAX_2	$a_{UU}\ a_{UU}\ d_{DS}\ c_{SD}\ b_{US}\ a_{DD}$	[32]
SAX Trend Distance (SAX-TD)	$a_{1.11}\ a_{0.62}\ d_{0.41}\ c_{0.82}\ b_{1.0}\ a_{0.55}$	[24]

the relation between the first and the second parts of the partition, and the second indicator represents the relation of the second and the third parts. The trend type can be up (U), down (D), and straight (S), and they are calculated the same way as in [14]. The authors also defined the corresponding distance measure, but they did not prove that it has the bounding-from-below property.

SAX-TD [25] also dealt with the missing trend information in SAX. In the proposed method, a new trend distance measure was introduced, which has tighter lower bound to Euclidean distance than that of the original SAX.

The 1d-SAX method, introduced in [23], is also a modified version of the SAX method. In 1d-SAX, a symbol encapsulates the mean and the trend of the partitions. The symbols are represented by bit strings in which the first part describes the mean value and the second describes the trend value. To determine the trend value, the 1d-SAX method fits linear regression to the data points of the partition. The performance analysis showed that the 1d-SAX method improves the speed of searching in large databases, while the dimensionality reduction ratio remained the same as in the original SAX method.

In [22], SAX is extended by calculating the maximum and minimum values of each partition in addition to the mean value. Therefore, each partition is represented by three symbols (symbol of min, max, and mean). The discretization step is the same as in SAX. The ordering of the three symbols, supposing the mean value is always in the middle of the partition, follows their original position in the time series. This extension enables to use the same $MINDIST$ distance measure, and thus it also has the lower bounding property.

Zan and Yamana [24] use the standard deviation to improve SAX. According to the authors, when comparing the spread of different time series that have approximately the same mean value, standard deviation provides the most valuable information for a Gaussian distribution. In their representation, the standard deviation of a partition is attached to the partition's symbol. With the standard deviation, the distance measure is improved and the time series classification also becomes more precise while the lower bounding property still exists.

The other limitation of SAX is that the quality of the symbolic representation highly depends on the right parameter selection. The authors of SAX applied a brute force method, an exhaustive search of the optimal parameter setting (summarized in Sect. 2.5), but this approach could be computationally intensive.

In [33], the authors proposed AutoiSAX, which estimates the SAX parameters based on the complexity of time series data and the change of the standard deviation between partitions. Complexity Estimate (CE) is used to estimate the partition size, K, based on the time series complexity, while the standard deviation among partitions is used to estimate the alphabet size, $|\mathcal{A}|$, which results in distinct $|\mathcal{A}|$ alphabet sizes for the time series in the dataset. Although their goal, obtaining the SAX parameters, was achieved, the CE does not work well for capturing the complexity of diverse time series data [27].

In DynamicSAX [27], the K partition size is estimated based on the Shannon sampling theorem and adaptive hierarchical segmentation, and the $|\mathcal{A}|$ alphabet size is set based on the skewness of the experimental distribution of PAA transformed time series.

The experimental results in [27] showed that the Shannon sampling based partition selection outperforms the AutoiSAX in parameter setting; however, the algorithm does not take into account the memory limitation, mentioned in Sect. 2.5.

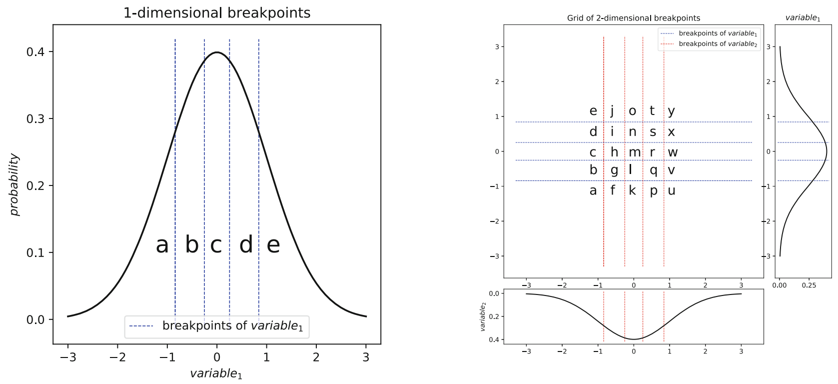
3 The Multivariate Symbolic Aggregate Approximation

As presented in Sect. 2, the literature is mainly focusing on the performance improvement of univariate time series representations, while in many applications multivariate time series are available. Multivariate time series describe the same process by more variables, and thus integrating the information from these variables in one symbol sequence has better expressive power than any other univariate solution. With a multidimensional distance function, which bounds from below the Euclidean distance, a multivariate representation could be extremely useful for implementing various data mining tasks like classification, pattern recognition, anomaly detection, or searching, with today's huge time series databases.

In this section, we are focusing on the introduction of our MSAX representation, which is able to represent multivariate time series in one sequence of symbols. During the construction of MSAX, we have kept in mind to preserve the two essential properties of SAX: equiprobability and distance measurement.

Let the $\boldsymbol{X}^{V \times N}$ matrix denote a multivariate time series, where V is the number of the variables and N is the number of the samples. The main idea behind our representation is that we consider each of $1, \ldots, V$ variables as a "dimension" in a V-dimensional space. Our goal is to transform each data sample from the V-dimensional space to a single symbol which represents a subset of the V-dimensional space separated by the breakpoints of the variables.

As an example, Fig. 1a shows the simplest 1-dimensional case, when the symbols split the real axis into segments, which satisfy the equiprobability property. This 1-dimensional case is identical with the original SAX. Figure 1b depicts a multivariate time series having two variables. In this case, the horizontal axis represents variable 1 and the vertical axis variable 2. The breakpoints of the two variables form a grid over the 2-dimensional space, where each cell has an associated symbol of the alphabet.



(a) 1-dimensional case, where the alphabet ($a = 5$) split the 1-dimensional space to 5 parts.

(b) 2-dimensional case, where two variables split the space to 25 cells. Each cell represents a symbol.

Fig. 1 1-dimensional and 2-dimensional examples of Multivariate Symbolic Aggregate Approximation (MSAX). The 1-dimensional case is identical with SAX. (**a**) 1-dimensional case, where the alphabet ($a = 5$) splits the 1-dimensional space to 5 parts. (**b**) 2-dimensional case, where two variables split the space to 25 cells. Each cell represents a symbol

In the 1-dimensional case, the equiprobability property means that all symbols from the used alphabet have the same probability due to the equally probable areas defined by the breakpoints (see Sect. 2.3). To preserve this property in the V-dimensional space, the variables have to be pairwise independent. In other words, each time series from the multivariate time series has to be uncorrelated from each other.

Let $\bar{\mathcal{V}} = \left[\mathcal{V}_1\ \mathcal{V}_2 \cdots \mathcal{V}_V\right]^T$ denote the vector of multivariate normal distributed random variable, in which every $\mathcal{V}_v$ random variable represents a time series, and ^{T} stands for the transpose. To check the independence of two $\mathcal{V}_f$ and $\mathcal{V}_l$ random variables ($f \neq l$), the Pearson correlation method can be used, defined by

$$Corr_{\mathcal{V}_f,\mathcal{V}_l} = \frac{E[(\mathcal{V}_f - \mu_{\mathcal{V}_f})(\mathcal{V}_l - \mu_{\mathcal{V}_l})]}{\sigma_{\mathcal{V}_f}\sigma_{\mathcal{V}_l}}, \tag{11}$$

where $\mu_{\mathcal{V}_f}$ and $\mu_{\mathcal{V}_l}$ are the expected values, and $\sigma_{\mathcal{V}_f}$ and $\sigma_{\mathcal{V}_l}$ are the standard deviations of the two variables obtained according to (1) and (2). If the absolute value of $Corr_{\mathcal{V}_f,\mathcal{V}_l}$ is under a certain ε threshold, $\mathcal{V}_f$ and $\mathcal{V}_l$ are considered to be independent.

In most real world applications, time series are not independent. Using Singular Value Decomposition (SVD), it is possible to introduce a linear transformation which converts the dependent variables to independent ones.

The covariance matrix of $\bar{\mathcal{V}}$ is

$$R_{\bar{\mathcal{V}}} = E[(\bar{\mathcal{V}} - \mu_{\bar{\mathcal{V}}})(\bar{\mathcal{V}} - \mu_{\bar{\mathcal{V}}})^T], \tag{12}$$

which is a real symmetric positive semidefinite matrix. The SVD of $R_{\bar{\mathcal{V}}}$ is

$$SVD(R_{\bar{\mathcal{V}}}) = \underbrace{\Theta^T \Lambda \Theta}_{SVD}, \tag{13}$$

where Θ is real unitary matrices, $\Theta^T \Theta = \Theta \Theta^T = I$, and Λ is a real diagonal matrix with non-negative diagonal entries.

Using the Θ and Λ matrices, we apply the following linear transformation on $\bar{\mathcal{V}}$:

$$\bar{\mathcal{Z}} = \Lambda^{-\frac{1}{2}} \Theta (\bar{\mathcal{V}} - \mu_{\bar{\mathcal{V}}}), \tag{14}$$

where $\bar{\mathcal{Z}}$ is the vector of transformed random variables $\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_V$. The transformed random variables are pairwise independent because

$$R_{\bar{\mathcal{Z}}} = E[(\bar{\mathcal{Z}} - \mu_{\bar{\mathcal{Z}}})(\bar{\mathcal{Z}} - \mu_{\bar{\mathcal{Z}}})^T] = \Lambda^{-\frac{1}{2}} \Theta \underbrace{E[\bar{\mathcal{V}}\bar{\mathcal{V}}^T]}_{\Theta^T \Lambda \Theta} \Theta^T \Lambda^{-\frac{1}{2}} = I, \tag{15}$$

where I is the identity matrix, which means that arbitrary $\mathcal{Z}_f$ random variable from $\bar{\mathcal{Z}}$ is independent of any other $\mathcal{Z}_l$ random variable if $f \neq l$ and the variance of every transformed $\mathcal{Z}_i$ is 1. The transformation also ensures that the mean of every transformed $\mathcal{Z}_i$ random variable is zero because

$$\mu_{\bar{\mathcal{Z}}} = E[\bar{\mathcal{Z}}] = E[\Lambda^{-\frac{1}{2}} \Theta (\bar{\mathcal{V}} - \mu_{\bar{\mathcal{V}}})] = \Lambda^{-\frac{1}{2}} \Theta E[\bar{\mathcal{V}} - \mu_{\bar{\mathcal{V}}}] = 0. \tag{16}$$

That is, the transformed $\bar{\mathcal{Z}}$ vector is composed of independent standard normal distributed random variables.

3.1 Steps of MSAX

The MSAX transformation includes four major steps; see Fig. 2.

3.1.1 Dependency Analysis

Dependency analysis is required to decide whether the linear transformation according to (14) is necessary or not. Given $\boldsymbol{X}^{V \times N}$ matrix, where V is the number of the time series, N is the length of time series and $X_{v,n}$ is the nth sample from vth time series. Using Eq. (11), the correlation coefficient between all time series can be calculated. If the absolute value of all correlation coefficient is under a sufficiently small ε, the time series of $\boldsymbol{X}^{V \times N}$ can be regarded as pairwise independent. In this case, the linear transformation is not needed, and therefore the next step is the

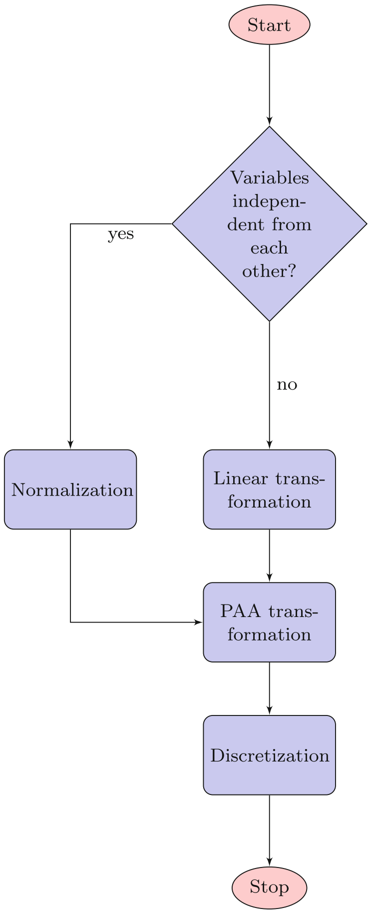


Fig. 2 Flow chart of MSAX transformation

normalization of time series, as stated in (3); otherwise, the linear transformation according to (14) is inevitable.

3.1.2 Linear Transformation

In the linear transformation step, (14) is applied on the $\boldsymbol{X}^{V\times N}$ matrix. The resulting matrix will be $\boldsymbol{Z}^{V\times N}$, whose v, n element, $Z_{v,n}$, is the nth sample from the vth transformed time series. Each row (time series) of $\boldsymbol{Z}^{V\times N}$ contains independent and normalized samples.

3.1.3 PAA Transformation

In the PAA transformation step, we introduce the w window size parameter, which denotes the maximal size of a partition in PAA. We use the w instead of K because the use of w accelerates the execution of the implementation. The K parameter, which is necessary to execute the PAA transformation (Sect. 2.2), is obtained as

$$K = \left\lceil \frac{N}{w} \right\rceil, \tag{17}$$

where N is the length of the time series and w is the window size.

The resulting $\check{\boldsymbol{Z}}^{V \times K}$ matrix ($K << N$) will contain the PAA transformed time series, in which $\check{Z}_{v,k}$ is the kth sample from the vth PAA transformed time series.

3.1.4 Discretization Step

In the discretization step, every $\check{\boldsymbol{Z}}_{v,k}$ PAA transformed value is translated to symbols. While in SAX and its extensions, a predefined table is used for breakpoints' calculation, we have defined a function that can determine the breakpoints values depending on the alphabet size $|\mathcal{A}|$. Using the inverse of the cumulative distribution function of normal distribution, the quantile function of the normal distribution can be constructed:

$$\Phi^{-1}(p) = \mu + \sigma\sqrt{2}\, erf^{-1}(2p - 1), \qquad p \in (0, 1). \tag{18}$$

To ensure the equiprobability, the probability of each symbol from alphabet $\mathcal{A}$ has to be $\frac{1}{|\mathcal{A}|}$. Modifying (18), for $a = 1, \ldots, |\mathcal{A}| - 1$, the breakpoints can be calculated in the following way:

$$\beta_a = \mu + \sigma\sqrt{2}\, erf^{-1}\left(2\frac{a}{|\mathcal{A}|} - 1\right) \overset{\mu=0,\ \sigma=1}{=} \sqrt{2}\, erf^{-1}\left(2\frac{a}{|\mathcal{A}|} - 1\right), \tag{19}$$

where erf^{-1} is the inverse error function. Using the SAX discretization (Sect. 2.3) combined with (19), the symbol sequence for each time series can be calculated based on the $|\mathcal{A}|$ alphabet size, independent of the alphabet itself. These operations produce the symbols in $\tilde{\boldsymbol{Z}}^{V \times K}$.

In the last step, the symbols of matrix $\tilde{\boldsymbol{Z}}^{V \times K}$ are transformed to a final symbol series $\tilde{\boldsymbol{Z}}$ with length K. To shrink the matrix in one sequence, the columns of the symbol matrix $\tilde{\boldsymbol{Z}}^{V \times K}$ are interpreted as new symbols from a V-dimensional space. The symbols in the kth column of $\tilde{\boldsymbol{Z}}^{N \times K}$ will be the coordinates of the symbol $\tilde{Z}_k$ ($k = 1, \ldots, K$). This means that a matrix $\tilde{\boldsymbol{Z}}^{V \times K}$ results in a symbol series $\tilde{\boldsymbol{Z}}$ with length K, while a symbol $\tilde{Z}_k$ compresses one column of the matrix. In Fig. 3, a 2-

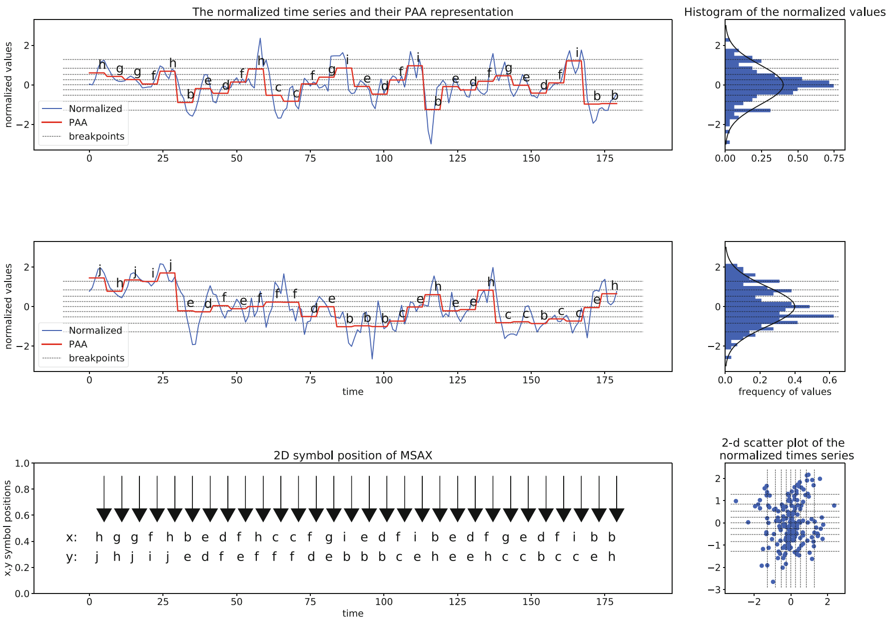


Fig. 3 An example of MSAX representation when the multivariate time series has 2 variables. The first two subfigures show the normalized and PAA transformed values of the 2 variables, together with the symbols and breakpoints. The alphabet size is set to $|\mathcal{A}| = 10$, while the window size is $w = 6$. The generated 1-dimensional symbols are placed over the corresponding windows. In the third subfigure, the 2-dimensional coordinates (x, y) of the MSAX representation are depicted

dimensional example is depicted. The first two subfigures plot the two time series with their symbol series with length 18, which form the symbol matrix $\tilde{\mathbf{Z}}^{2\times 18}$. In the third subfigure, the 2-dimensional coordinates (x, y) of the MSAX representation are depicted. For instance, the coordinates of first symbol of MSAX representation are (h, j).

4 The MSAX Distance

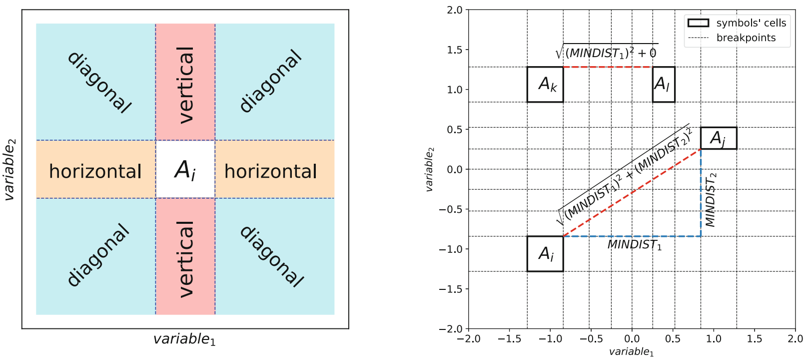
Besides the equiprobability, another important feature of the original SAX is the possibility of distance measurement. Because of the PAA and the discretization steps, there is an information loss which cannot be restored from the transformed time series; thus, the Euclidean or other well-known distance function cannot be used. However, the authors of the original SAX defined the $MINDIST$ distance measure being able to calculate a proven bounding-from-below distance of the Euclidean distance. The tightness of the lower bound depends on the w window size and $|\mathcal{A}|$ alphabet size. Some extensions of PAA do not provide the possibility of distance calculation, while this property is crucial for MSAX usability. Without the distance measure, clustering algorithms and some classification algorithms cannot be used [34].

In this section, we define a new distance function that integrates the original $MINDIST$ function, and thus it keeps the important property that bounds from below the Euclidean distance.

The MSAX distance measurement is the extended version of the $MINDIST$, and it reduces the V-dimensional distance calculation to a 1-dimensional one. At first, let us examine the 2-dimensional case, where we have a Euclidean space in which the axioms and postulates of the Euclidean geometry can be applied. In this space, the breakpoints of the two dimensions split the space to an irregular grid.

Let us consider two arbitrary symbols, A_i and A_j, and determine the distance between them. Obviously, if $A_i = A_j$, then the distance will be zero. If $A_i \neq A_j$, then eight different situations can be distinguished which can be categorized in the following two groups:

1. A_j is in a horizontal or a vertical direction to A_i (depicted in Fig. 4a) and
2. A_j is in a diagonal direction to A_i (depicted in Fig. 4a).



(a) The three distinct cases (diagonal, vertical and horizontal) we have to deal during the distance calculation.

(b) In the diagonal case, the Euclidean distance of $MINDIST$s will give the distance between two symbols ($A_i \leftrightarrow A_j$). This method is sufficiently universal to calculate horizontal or vertical distances as well ($A_k \leftrightarrow A_l$).

Fig. 4 Visualization of M_DIST distance calculation. (**a**) The three distinct cases (diagonal, vertical, and horizontal) we have to deal during the distance calculation. (**b**) In the diagonal case, the Euclidean distance of $MINDIST$s will give the distance between two symbols ($A_i \leftrightarrow A_j$). This method is sufficiently universal to calculate horizontal or vertical distances as well ($A_k \leftrightarrow A_l$)

4.1 Horizontal and Vertical Cases

In the horizontal and vertical cases, the two symbols have the same position along one of the dimensions. Since the same position means zero $MINDIST$, only the other dimension's $MINDIST$ has to be taken into account which lower bounds the Euclidean distance. This idea is also valid with more dimensions; thus, each dimension where the symbols' positions are the same can be left out from the distance calculation.

4.2 Diagonal Case

The more interesting and more general case is the diagonal one, where none of the two symbols' positions are the same along any dimension. In two dimensions, each symbol has its own rectangular cell; see Fig. 4a or b. Because of the irregular grid of breakpoints, we can be sure that the length of the shortest path between two symbols (or rectangular cells) is the distance between the nearest corners of the cells.

Using that observation, we just need to calculate the Euclidean distance of $MINDIST$s. As you can see in Fig. 4b, the $MINDIST$ for each dimension has to be determined at first, and then the Euclidean distance can be calculated based on the previous values. This method works similarly in V dimensions:

$$M\_DIST = \sqrt{\sum_{v=1}^{V}(MINDIST_v)^2}, \tag{20}$$

where N is the number of the dimensions. Equation (20) is sufficiently universal to cover the horizontal and vertical cases (see Fig. 4b), and because the M_DIST uses the $MINDIST$, it also inherits the lower bounding property.

5 MSAX Parameter Optimization

As we mentioned in Sect. 1, the fine-tuning of SAX parameters is essential to achieve the right symbolic representation. In this section, as the MSAX representation is highly dependent on the original SAX, we are focusing on the appropriate setting of $|\mathcal{A}|$ and w parameters of SAX. In our solution, a new objective function has been developed, which supports the utilization of different derivate-free optimization algorithms. These algorithms converge to the global optimal setting for SAX and MSAX but are also suitable for the parameter setting of several extensions of SAX.

To find the best setting, the authors of [21] empirically determined the best values by defining the tightness of lower bound (9). Unfortunately, this measure does not support different optimization techniques because it uses two time series in the tightness calculation instead of one. In case of a dataset which contains M time series, this approach has to execute $M^2 - M$ calculations, which makes the empirical parameter search much more computationally intense than it should be.

Our solution differs from (9) as it examines the dissimilarity between the original raw time series and the time series' SAX transformed version. Dissimilarity (or error) comes from two sources: (1) PAA transformation and (2) symbolic transformation.

The error of the PAA transformation (4) with a certain w window size comes from the difference between the mean value and the real value in the certain windows. By increasing the w window size, the error gets bigger. For calculating this error, the following equation can be used:

$$paa\_err(\boldsymbol{X}, w) = \sum_{k=0}^{K-1} \sum_{n=kw+1}^{w(1+k)} \left\| X_n - \check{X}_{k+1} \right\|_1, \tag{21}$$

where K is the length of the $\check{\boldsymbol{X}}$ time series, w is the currently used window size, $\boldsymbol{X}$ is the original time series, $\check{\boldsymbol{X}}$ is the PAA transformed version, and $|| \, ||_1$ is the L_1 normal distance.

The error of the symbolic transformation comes from the quantization. Each symbol A_a has an upper bounding β_{a+1} and a lower bounding β_a breakpoint (except the lowest and highest symbol). To calculate the error, each PAA transformed value is compared to the mean of these two breakpoints:

$$sym\_err(\boldsymbol{X}, |\mathcal{A}|) = w \sum_{k=1}^{K} \left\| \check{X}_k - \frac{u(\tilde{X}_k) + l(\tilde{X}_k)}{2} \right\|_1, \tag{22}$$

where K is the length of the $\check{\boldsymbol{X}}$ time series, w is the window size, $\boldsymbol{X}$ is the original time series, $\check{\boldsymbol{X}}$ is the PAA transformed version, $\tilde{\boldsymbol{X}}$ is the result of MSAX, $u(.)$ function returns the upper bounding breakpoint of a symbol, $l(.)$ function returns the lower bounding breakpoint of a symbol, and $|| \, ||_1$ is the L_1 normal distance.

Utilizing the paa_err and sym_err error functions, the optimization problem of w and a parameters can be described by the following equation:

$$\begin{aligned} |\mathcal{A}|, w :&= \arg\min_{|\mathcal{A}|, w} \{J(\boldsymbol{X}, |\mathcal{A}|, w)\} \\ &= \arg\min_{|\mathcal{A}|, w} \{paa\_err(\boldsymbol{X}, w) + sym\_err(\boldsymbol{X}, |\mathcal{A}|)\}. \end{aligned} \tag{23}$$

Because we did not set an upper limit for $|\mathcal{A}|$ alphabet size, we had got extremely high $|\mathcal{A}|$ values (1000+) in the first executions of the optimization algorithms. In real world scenarios, these extreme parameter values are pointless, so a simple L_1 regularization had been added to the error function:

$$J(\boldsymbol{X}, |\mathcal{A}|, w)\} = paa\_err(\boldsymbol{X}, w) + sym\_err(\boldsymbol{X}, |\mathcal{A}|) + \underbrace{\lambda(|\mathcal{A}| + w)}_{L_1 regularization}, \tag{24}$$

where λ is a parameter which controls the importance of the regularization term.

In the original paper [21], the authors mentioned that if we wish to approximate a massive dataset in main memory, memory limitation also has to be taken into account during the parameter setting.

A symbol can be represented by a $log_2(|\mathcal{A}|)$ long binary string if the alphabet size is $|\mathcal{A}|$. If N is the length of the time series and w is the window size, the $\frac{N}{w}$ fraction determines the number of required symbols for the symbolic representation. On the basis of the precedings, the *mem* memory limit of the main memory modifies Eq. (23) as

$$|\mathcal{A}|, w := \underset{|\mathcal{A}|, w: \frac{N}{w} log_2(|\mathcal{A}|)}{\arg\min} \{J(\boldsymbol{X}, |\mathcal{A}|, w)\}. \tag{25}$$

We have visualized the error surface to understand how the error behaves for different combinations of the parameter values and how the *mem* changes the feasible parameter range. For the visualization, all possible $|\mathcal{A}|$ and w combinations have been calculated, for the $|\mathcal{A}|$'s value range of [3, 200] and w's value range of [2, 50]. This method is identical to the original parameter setting method.

As data source, datasets from [35] were used. During the error surface calculation, the different datasets gave distinct surfaces, and however their shapes were always similar to Fig. 5.

Examining Fig. 5, the effect of the parameters' changing is clearly observable. In case of small alphabet size values, the error is prominently high, but the increase of alphabet size results in a hyperbolic decay until it starts to increase because of the L_1 regularization. The effect of window size is quite the opposite. Higher window sizes cause higher errors, and therefore the window size should be kept low. Another interesting observation is the saw-toothed edges of the surface. This is due to the memory limitation because the parameter pairs that have higher memory demands than the current limit cannot been calculated. Thus, when selecting the parameters, one should tend to choose the smallest w and the biggest $|\mathcal{A}|$ combination available, while not exceeding the memory limit.

To reveal the invisible parts of the surface, the available memory size has to be increased. In Fig. 6, the effect of memory size is depicted. While the available memory size is increased, the edges become smoother and smoother, and better parameter combinations are to be found, since the larger memory allows more accurate symbolic representation.

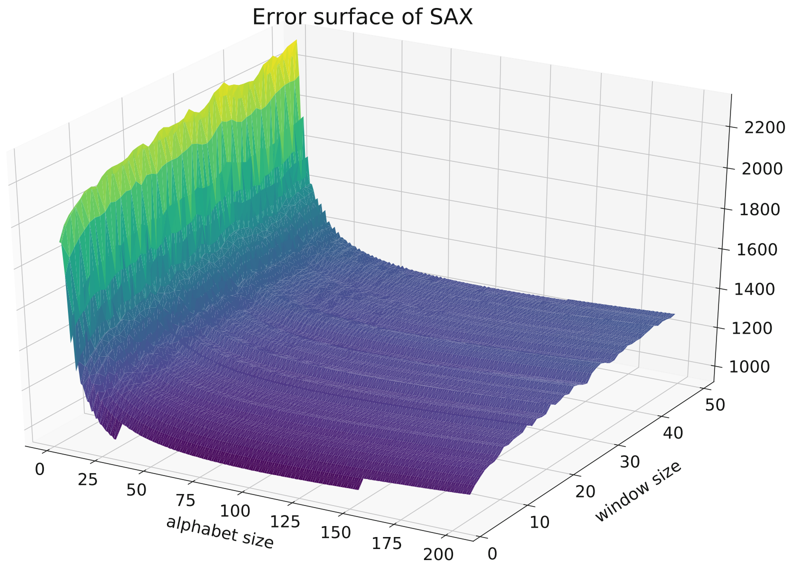


Fig. 5 An example error surface

In the ideal case when enough memory is provided, the w window size parameter can be set to 2, which is the lowest (and the best) interpretable value. Replacing w with 2 in (25), the optimal $|\mathcal{A}|$ alphabet size can be calculated based on the provided high memory mem:

$$\frac{N}{2} log_2(|\mathcal{A}|) = mem \quad \Leftrightarrow \quad |\mathcal{A}| = 2^{\frac{2\, mem}{N}}. \tag{26}$$

Unfortunately, determining the $|\mathcal{A}|$ alphabet size with (26) is not feasible in most of the cases, because the memory limitation does not allow the usage of $w = 2$ parameter setting. Thus, other methods have to be used which can deal with the memory limitation.

Since the defined error function is not continuous because of the steps of symbolic transformation, the usage of well-known gradient descent optimization procedures is not possible. Therefore, to optimize our error function, we have used different derivate-free optimization algorithms such as Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and Particle Swarm Optimization (PSO).

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [36] is a stochastic derivative-free numerical optimization algorithm for difficult (non-convex, ill-conditioned, multi-modal, noisy) optimization problems. In this evolution strategy, new candidate solutions are sampled according to a multivariate normal distribution. Taking the results of each generation, it adaptively increases or decreases the search space for the next generation by changing the covariance matrix of the multivariate normal distribution. BI-population CMA-ES (BIPOP-CMA-ES) [37] is a multistart

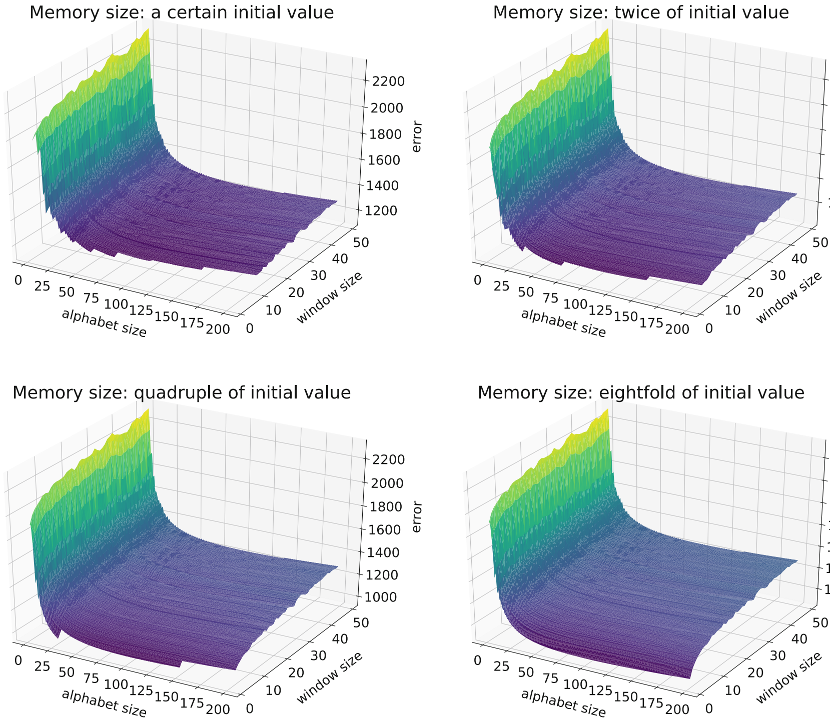
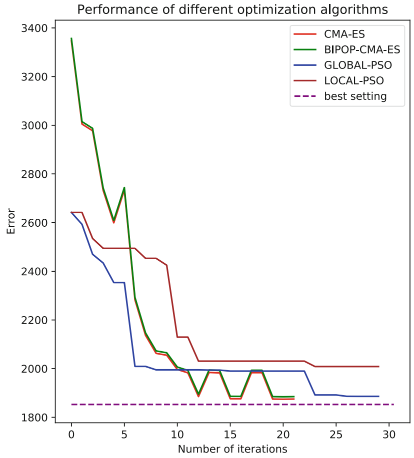


Fig. 6 The effect of memory size

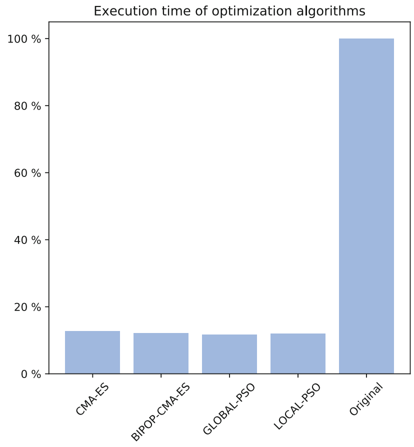
extension of CMA-ES; one starts with an increasing population size and another with varying small population sizes.

Particle Swarm Optimization [38] is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality (error function). The idea of Particle Swarm Optimization (PSO) is to emulate the social behavior of birds and fishes by initializing a set of candidate solutions to search for an optima. Particles are scattered around the search space, and they move around it to find the position of the optima. Each particle represents a candidate solution, and their movements are affected in a twofold manner: their cognitive desire to search individually (local) and the collective action of the group or its neighbors (global).

Our performance analysis has two purposes. First, we aim to find the optimization techniques that give the best result with our error function and also being faster than the original one [21]. To compare the performance of the different optimization methods on the datasets and check the correctness of our error function, we use the [39] PSO implementation and [40] CMA-ES implementation. On the other hand, our goal is to check that our error function truly results in a small w and big a combination.



(a) Error decreasing performance of the derivate free optimization methods



(b) Execution time of the applied optimization methods.

Fig. 7 Comparison of optimization algorithms' performance. (**a**) Error decreasing performance of the derivate-free optimization methods. (**b**) Execution time of the applied optimization methods

The results are depicted in Fig. 7a; however, they always depend on the current dataset and the right hyperparameter setting, and thus the selection of a single best optimization method is not possible.

Our performance analysis indicates that almost every optimization algorithm converges to a near-optimal solution (the optimum is marked by the purple dashed line in the figure), except the Local-PSO whose particles stuck in the local minimums of the error surface. It can be seen that the CMA-ES and BIPOP-CMA-ES converges faster to the best solution than the two PSO methods. It can be also observed that CMA-ES and BIPOP-CMA-ES give the same results, so we recommend to use CMA-ES because it has shorter execution time.

The execution time of the algorithms is also compared to the original (empirical) one [21]; see Fig. 7b. As you can see, the methods outperform the original, but a significant difference between them is not observable on the examined dataset. The optimization algorithms achieved almost a tenfold speedup over the original, with our error function.

6 MSAX Performance Evaluation

In this section, we will present the results of the performance evaluation of our MSAX method. At first, the used datasets are introduced, and then the methodology of the evaluation is discussed. For the sake of reproducibility of evaluation results, we made our implementation publicly available [41].

6.1 Datasets

The used datasets are provided by the UCI Time Series repository [35]. Since our goal is to develop a new method which improves the SAX performance in case of multivariate datasets, we chose 9 multivariate smart city-related datasets. During the selection of the types of the datasets, we focus on real Human Activity Recognition (HAR) datasets, as they can provide significant information for the optimization of smart city services (like daily activity classification). They have from 3 to 14 variables. Each dataset is divided into a training set and a testing set which are 80 and 20%, respectively. The chosen datasets contain classes ranging from 2 to 22 and are of size from dozens to thousands. The classes are the possible categories to be predicted. For example, a HAR dataset typically contains *standing*, *walking*, *going up/down stairs*, etc. classes.

6.2 Methodology

For the performance evaluation of our method, we apply a classification task on the datasets, where the classification accuracy measure is the basis of the comparison. Since the First Nearest Neighbor (1-NN) algorithm is the standard time series classification method for comparison in the literature, we also use it during the evaluations. In general, 1-NN will not give the best classification result, but the performance of symbolic representations can be easily measured.

Besides the classification accuracy, the other important performance metric is the Dimensionality Reduction Ratio (DRR) [25], which can express the compactness of a symbol sequence compared to the original. DRR can be calculated using (27).

$$\text{DRR} = \frac{\text{Number of reduced data points}}{\text{Number of original data points}} \tag{27}$$

In the analysis, Multivariate Symbolic Aggregate Approximation (MSAX) is compared to SAX. For the sake of fairness, MSAX and also SAX use our M_DIST distance measure too, which makes possible to calculate the distance in case of multiple variables (as the original SAX cannot handle multivariate time series).

During the performance analysis, the best $|\mathcal{A}|$ and w parameters for each dataset are determined by the CMA-ES optimization method using our objective function with $N = 5000$ and $mem = 10000$ bits. In large proportion of real life applications, it is crucial to get fast feedback from the classification method. Thus, instead of comparing the whole time series, the bag-of-words [42] method is used in which the whole symbol sequence is split into smaller subsequences that incorporate the words. We have to highlight that the right word size always depends on the length of the patterns in the investigated dataset. Therefore, we use the same word size for all datasets because our goal was the comparison of the two approximation methods.

Based on our experiments, we set the $word\_size$ to 30, because it is long enough to capture the essential patterns in general.

6.3 Results

The overall classification results are listed in Table 2, where entries with the highest classification accuracies and the lowest DRR are in bold.

Since one of the important features of the SAX representation is its dimensionality or numerosity reduction, we have compared the dimensionality reduction of our method with the SAX. The result of the dimensionality reduction basically depends on the value of w window size. In our experiments, the value of w was typically under 4, since the optimization process tries to find the most accurate parameter setting, which yields the low w values.

The MSAX method always achieved better DRR results compared to SAX in the case of the investigated datasets. It is due to the fact that MSAX can highly increase the express power of one symbol depending on the number of variables of the data. Based on Table 2, MSAX can reduce the number of data points to under 4% of the original number of data points in some cases, while the classification accuracy is between 90 and 100% using 1-NN. In average, MSAX decreases the DRR by 31% compared to SAX.

Examination of classification results shows that MSAX performs better in the majority of the investigated datasets, even though SAX and MSAX used the same distance measurement. From the nine datasets, MSAX surpasses SAX in seven cases. In the other two cases, the difference between the two methods' accuracy was minimal. The reason behind the good results is the Independency Transformation, which removes the noise parts between the variables. The MSAX improved the classification accuracy by 22% in average when compared to the results of the SAX.

6.4 Execution Time of MSAX

Another important criterion of the usability of the MSAX is its execution time. To examine the performance of MSAX from this aspect, we need to check the effect of the growth of the number of variables (dimensions). For the evaluation, the EEG dataset was chosen because it has the highest number of variables (14 distinct variables and 209,720 data points) from the investigated datasets.

We start the evaluation with the first variable from the dataset. In every forthcoming step, we add the next variable to the set of analyzed variables until we reached the 14 variables overall. In each step, the execution times of MSAX and the SAX are measured. In the case of MSAX, it was enough to execute the method only once, while for the SAX it had to be applied for the current variables separately, because it can handle only one variable at a time.

Table 2 1-NN classification accuracy; the lowest error rates and the highest DRRs are highlighted in bold

No.	Dataset name	# Variables	# Classes	# Data points	Alphabet size	Window size	SAX		MSAX	
							Accuracy	DRR	Accuracy	DRR
1	HAR	3	7	7,692,708	63	3	0.3856	0.3333	**0.6916**	**0.1111**
2	UWI	3	22	593,684	17	2	0.3063	0.5	**0.6486**	**0.1667**
3	AREM	6	7	295,673	23	2	0.315	0.5	**0.5909**	**0.0833**
4	HAPT	3	12	5,709,298	29	2	**0.3389**	0.5	0.2748	**0.0833**
5	DAPHNET	9	2	11,197,140	64	3	0.8961	0.3333	**0.9**	**0.037**
6	EEG	14	2	209,720	28	3	0.5532	0.3333	**0.8333**	**0.0417**
7	Occupancy	9	2	13,614,759	69	3	0.8356	0.3334	**1.0**	**0.0667**
8	EMG	8	7	123,360	67	3	**0.2922**	0.3333	0.275	**0.0417**
9	HMP	3	14	1,736,508	30	3	0.2556	0.3333	**0.4222**	**0.1111**

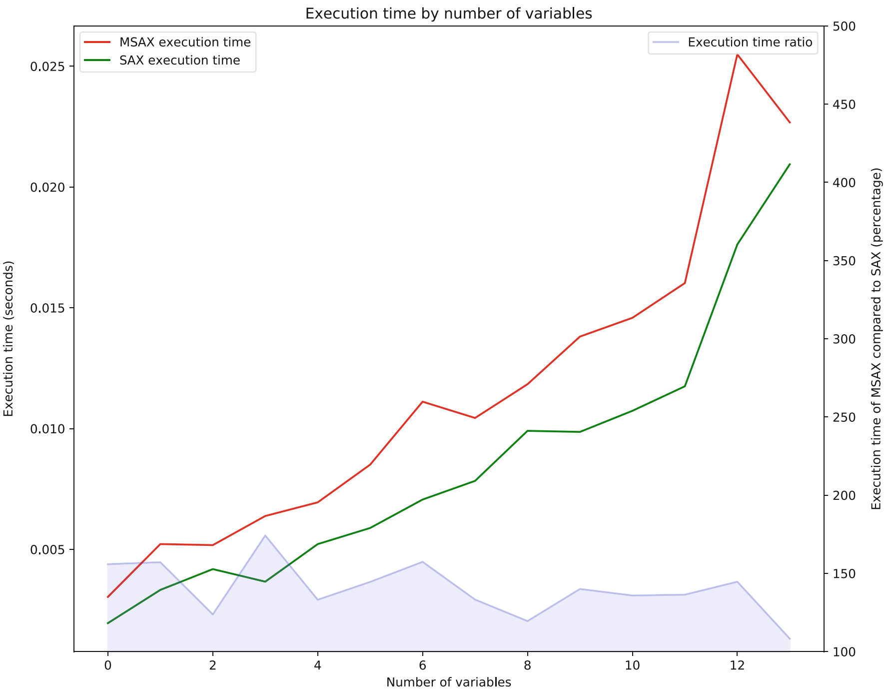


Fig. 8 The execution times of SAX and MSAX and their ratio

The execution times of SAX and MSAX are depicted in Fig. 8, where the red line represents the MSAX, while the green represents the SAX. The execution time increases linearly with the number of variables in both cases. As we expected, the execution time of MSAX is always higher because of additional computations like Independency Transformation (see Sect. 3.1.2). In the same Figure, the light purple area depicts the ratio of the two execution times ($\frac{MSAX}{SAX}$). As you can see, MSAX requires around 50% more time to calculate the symbolic representation, which is only 0.003 s in average. By raising the number of variables, the execution time ratio slightly decreases. Much more importantly, it increases the express power of the symbolic representation significantly, while the price is only a low additional cost in execution time.

7 Conclusion

We have proposed the Multivariate Symbolic Aggregate Approximation (MSAX) method as an extension of SAX, which allows to represent multivariate time series with only one sequence of symbols. MSAX representation can help the data scientist in the efficient and accurate processing of smart cities' enormous time series datasets.

At first, we introduced the idea behind our approach, and then the steps of MSAX were discussed in detail. The new method fulfills the equiprobability property. We defined also a new distance measurement which integrates the original $MINDIST$ function, and thus it keeps the important property that lower bounds the Euclidean distance. To find the best available parameter setting, we defined a new objective function, which makes the usage of different derivate-free optimization techniques possible.

Based on our experiments, MSAX significantly increased the accuracy of the standard 1-NN classification compared to SAX, while it could reduce the original dataset size to 96% in several cases.

References

1. S.h. An, B.H. Lee, D.R. Shin, in *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks* (IEEE, 2011), pp. 332–337
2. F. Ahmed, Y. Hawas, Transportation Research Part C: Emerging Technologies **60**, 52 (2015)
3. Caltrans performance measurement system (PEMS). URL http://pems.dot.ca.gov/
4. T. Lin, H. Rivano, F. Le Mouël, IEEE Transactions on Intelligent Transportation Systems **18**(12), 3229 (2017)
5. B. Siregar, A.B.A. Nasution, F. Fahmi, in *2016 International Conference on ICT For Smart Society (ICISS)* (IEEE, 2016), pp. 49–52
6. S. Seto, W. Zhang, Y. Zhou, in *2015 IEEE Symposium Series on Computational Intelligence* (IEEE, 2015), pp. 1399–1406
7. N. Dawar, N. Kehtarnavaz, in *2018 IEEE 14th International Conference on Control and Automation (ICCA)* (IEEE, 2018), pp. 482–485
8. A. Fleury, M. Vacher, N. Noury, IEEE transactions on information technology in biomedicine **14**(2), 274 (2009)
9. X. Su, H. Tong, P. Ji, Tsinghua science and technology **19**(3), 235 (2014)
10. A. Bagnall, J. Lines, A. Bostrom, J. Large, E. Keogh, Data Mining and Knowledge Discovery **31**(3), 606 (2017)
11. S. Aghabozorgi, A.S. Shirkhorshidi, T.Y. Wah, Information Systems **53**, 16 (2015)
12. H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.A. Muller, Data Mining and Knowledge Discovery pp. 1–47 (2019)
13. M. Ahmed, A.N. Mahmood, M.R. Islam, Future Generation Computer Systems **55**, 278 (2016)
14. B. Esmael, A. Arnaout, R.K. Fruhwirth, G. Thonhauser, in *International Conference on Computational Science and Its Applications* (Springer, 2012), pp. 392–403
15. R. Agrawal, C. Faloutsos, A. Swami, in *International conference on foundations of data organization and algorithms* (Springer, 1993), pp. 69–84
16. Y. Qu, B.l. Adam, M. Thornquist, J.D. Potter, M.L. Thompson, Y. Yasui, J. Davis, P.F. Schellhammer, L. Cazares, M. Clements, et al., Biometrics **59**(1), 143 (2003)
17. B.K. Yi, C. Faloutsos, in *VLDB*, vol. 385 (Citeseer, 2000), vol. 385, p. 99
18. C. Guo, H. Li, D. Pan, in *International conference on knowledge science, engineering and management* (Springer, 2010), pp. 234–244
19. E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, ACM Sigmod Record **30**(2), 151 (2001)
20. I.V. Oseledets, E.E. Tyrtyshnikov, SIAM Journal on Scientific Computing **31**(5), 3744 (2009)

21. J. Lin, E. Keogh, L. Wei, S. Lonardi, Data Mining and knowledge discovery **15**(2), 107 (2007)
22. B. Lkhagva, Y. Suzuki, K. Kawagoe, DEWS2006 4A-i8 **7** (2006)
23. S. Malinowski, T. Guyet, R. Quiniou, R. Tavenard, in *International Symposium on Intelligent Data Analysis* (Springer, 2013), pp. 273–284
24. C.T. Zan, H. Yamana, in *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services* (ACM, 2016), pp. 72–80
25. Y. Sun, J. Li, J. Liu, B. Sun, C. Chow, Neurocomputing **138**, 189 (2014)
26. A.M. Nagy, V. Simon, Pervasive and Mobile Computing **50**, 148 (2018). DOI https://doi.org/10.1016/j.pmcj.2018.07.004
27. C.T. Zan, H. Yamana, International Journal of Web Information Systems **13**(4), 387 (2017)
28. E. Keogh, S. Kasetty, Data Mining and knowledge discovery **7**(4), 349 (2003)
29. E. Keogh, S. Lonardi, B.c. Chiu, in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2002), pp. 550–556
30. R.J. Larsen, M.L. Marx, et al., *An introduction to mathematical statistics and its applications*, vol. 2 (Prentice-Hall Englewood Cliffs, NJ, 1986)
31. A. Apostolico, M.E. Bock, S. Lonardi, Journal of Computational Biology **10**(3–4), 283 (2003)
32. K. Zhang, Y. Li, Y. Chai, L. Huang, in *2018 Chinese Control And Decision Conference (CCDC)* (IEEE, 2018), pp. 2234–2240
33. N.C. Castro, P.J. Azevedo, Intelligent Data Analysis **19**(3), 581–595 (2015). URL https://doi.org/10.3233/IDA-150733
34. R. Xu, D.C. Wunsch, (2005)
35. D. Dua, C. Graff. UCI machine learning repository (2017). URL http://archive.ics.uci.edu/ml
36. N. Hansen, in *Towards a new evolutionary computation* (Springer, 2006), pp. 75–102
37. N. Hansen, in *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers* (ACM, 2009), pp. 2389–2396
38. R. Eberhart, J. Kennedy, in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science* (IEEE, 1995), pp. 39–43
39. L.J.V. Miranda, Journal of Open Source Software **3** (2018). URL https://doi.org/10.21105/joss.00433
40. N. Hansen, Y. Akimoto, P. Baudis. Cma-es/pycma on github (2019). URL https://doi.org/10.5281/zenodo.2559635
41. A.M. Nagy. Msax (2019). URL https://doi.org/10.5281/zenodo.3402973
42. P. Senin, S. Malinchik, in *2013 IEEE 13th international conference on data mining* (IEEE, 2013), pp. 1175–1180

A Pedestrian-Level Strategy to Minimize Outdoor Sunlight Exposure

Xiaojiang Li, Yuji Yoshimura, Wei Tu, and Carlo Ratti

1 Introduction

Too much sunlight exposure is thermally uncomfortable and potentially dangerous for people (Brash et al. 1991; Hodder and Parsons 2007; Kurazumi et al. 2013; Richards and Edwards 2017; Young 2009), although the minimum amount of sunlight is required for humans. Sunlight exposure is a major factor of human heat stress, which is a comprehensive parameter influenced by humidity and temperature (Gaffen and Ross 1998; NOAA 2009). Heat stress in summer poses a daily health threat to urban residents and causes morbidity and mortality (Stone et al. 2010). Other than heat stress, unprotected exposure to ultraviolet (UV) radiation in the sunlight is one of the major risk factors for skin cancers (Brash et al. 1991; Armstrong and Kricker 2001). Preventing too much sunlight exposure would help people to decrease the potential dangers caused by sunlight.

Urban streets that carry most of the human outdoor activities in cities (Li et al. 2017) are the major place for human outdoor sunlight exposure. Understanding the

X. Li (✉)
Department of Geography and Urban Studies, Temple University, Philadelphia, PA, USA

Y. Yoshimura
Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan

W. Tu
Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services and Research Institute of Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China

C. Ratti
Senseable City Lab, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_7

spatiotemporal distribution of sunlight in street canyons would help us to develop methods to protect people from too much outdoor sunlight exposure. The solar radiation reaching the street canyons is influenced by the obstruction effects of street trees and buildings (Carrasco-Hernandez et al. 2015; Hwang et al. 2011; Johansson 2006; Lin et al. 2010). Buildings and trees on both sides of streets would obstruct sunlight and then reduce the potential human sunlight exposure in street canyons (Li et al. 2018). The urban form and the orientation of the street canyons would influence the obstruction effects of buildings and tree canopies on the sunlight (Algeciras et al. 2016; Ali-Toudert and Mayer 2006; Johansson 2006). Usually, those streets with larger height-weight ratio have less incoming sunlight reaching the ground. Those east-west orientation streets have more solar radiation reaching the ground because of the same direction with the sun's zenith (Sanusi et al. 2016). In different times of 1 day, the different sun positions and the surrounding obstructions would make sunlight exposure significantly different.

The high-resolution building height models make it possible to simulate the obstruction effects of building blocks on sunlight and estimate the transmission of sunlight within street canyons, which further makes it possible to generate the spatiotemporal distributions of sunlight in the street canyons. However, building height models usually oversimplify the geometries of urban canyons (Carrasco-Hernandez et al. 2015). In addition, building height models only consider the shading effects of buildings and the shading effects of tree canopies and other urban features are usually not considered (Li et al. 2018).

The street-level images provide a new approach to simulate the sunlight within street canyons. Different from the building height models, which usually only include the information about building blocks, the street-level images can represent all types of obstructions along the streets (Li et al., 2018; Gong et al. 2018). Therefore, using street-level images would be more reasonable to simulate the shading effects of obstructions and generate more accurate spatiotemporal distributions of sunlight in street canyons. In addition, street-level images, such as Google Street View (GSV) images and Mapillary images, are globally available. All of these make street-level images suitable for predicting human outdoor sunlight exposure along streets.

In this study, we propose an individual-level strategy for people to minimize the outdoor sunlight exposure based on the simulation of the sunlight in street canyons using the GSV images as a surrogate of the streetscape environment. A pre-trained deep learning model (PSPNet model) was used to segment the street-level images and recognize the obstructions of sunlight within street canyons. We then generated the spatiotemporal distribution of sunlight exposure in street canyons by calculating the sun positions over times and projecting sun positions on the segmented GSV images. Based on the generated spatiotemporal distribution of sunlight exposure in the study area, we further developed a spatiotemporal routing algorithm to provide an individualized routing choice for people to minimize their sunlight exposure. The proposed individual-level strategy would help to reduce the negative effects of sunlight exposure for urban residents.

2 Study Area and Dataset

The study area, Shibuya, is a special ward and a major commercial and business center in Tokyo, Japan. Shibuya is very densely populated with an estimated population of 221,800 and population density of 14,679.09 people per km^2. In order to collect GSV images in the study area, we created sample sites along streets every 10 m (Fig. 1(a)). The street map used in this study was collected from Open Street Map. The coordinates of these sample sites were then used to collect the metadata (Fig. 1(b)). This study focuses on human sunlight exposure in hot summer; therefore, only those images captured in similar seasons were used in this study. Based on the time stamps in the collected metadata of GSV panorama, we only selected the most recently captured images in summer for each sample site and downloaded GSV panoramas (Fig. 1(c)).

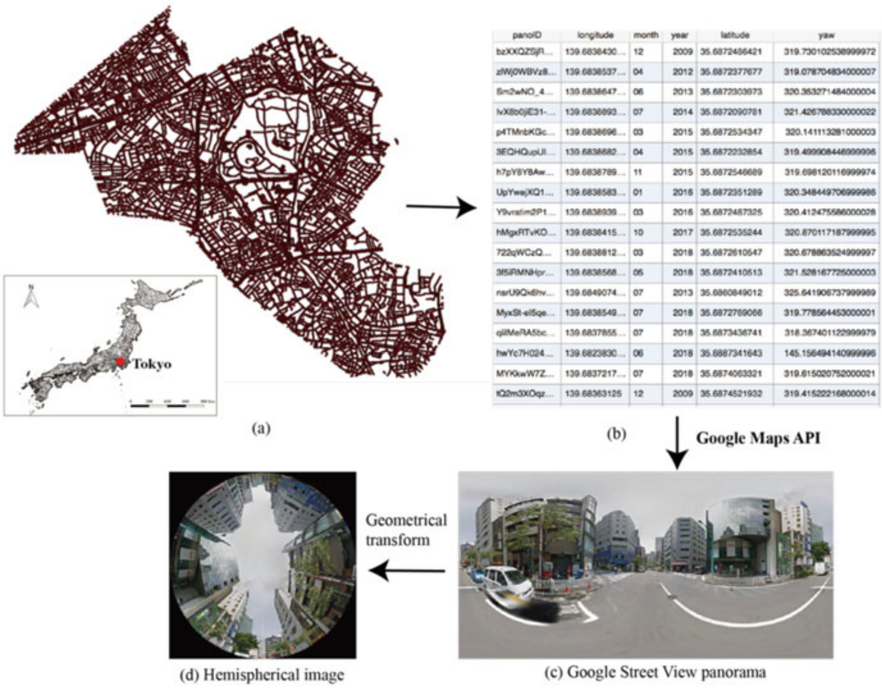


Fig. 1 The workflow for *GSV* google street view panorama collection in Shibuya, Tokyo, Japan, (**a**) the created sample sites in the study area, (**b**) the metadata of GSV panoramas, (**c**) a downloaded GSV panorama, (**d**) a generated hemispherical image based on geometrical transform

3 Methodology

3.1 Hemispherical Image Generation and Segmentation

Hemispherical image-based method is one of the standard methods for estimating solar radiation reaching the ground (Rich 1990; Easter and Spies 1994; Matzarakis et al. 2007, 2010). In this study, hemispherical images were generated from cylindrical GSV panoramas by geometrical transform (Li et al. 2017, 2018). In order to derive the sunlight obstruction information in street canyons, the image segmentation algorithm PSPNet was used to segment GSV panoramas into sky pixels and obstruction pixels (vegetation, buildings, and impervious surface pixels) (Gong et al. 2018; Zhao et al. 2017). The segmented GSV panoramas were further geometrically transformed into hemispherical images, which were used to model the human sunlight exposure within the street canyons. Figure 2 shows the segmentation results (Fig. 2(b)) of PSPNet on three GSV panoramas (Fig. 2(a)) and the segmented hemispherical images based on GSV panorama segmentation results (Fig. 2(c)).

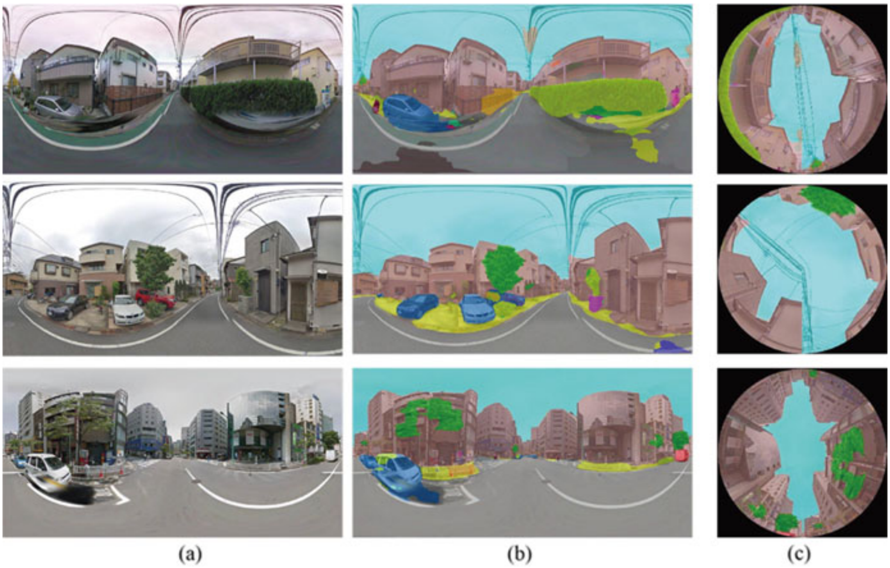


Fig. 2 Image segmentation results using PSPNet, (**a**) original Google Street View (GSV) panoramas, (**b**) the blend of segmentation results on GSV panoramas, (**c**) the hemispheric view of the segmented GSV panoramas

3.2 Human Exposure to Sunlight in Street Canyon

Human outdoor sunlight exposure is influenced by the time, orientation of the streets, buildings, and street tree canopies. Based on the generated segmented hemispherical images, it is possible to estimate whether a pedestrian is exposed to sunlight or not at any time and location by projecting the sun position on the hemispherical images.

In this study, we implemented the sun position estimation algorithm developed by NOAA Earth System Research Laboratory (ESRL, https://www.esrl.noaa.gov/) to estimate the sun position at any specific time. Figure 3 (a) depicts the geometrical model of the sun and a pedestrian. Figure 3 (b) shows the projected sun positions in 1 day on three hemispherical images in the study area. For a person at the location of (*lon*, *lat*), if the sun at one time (*t*) is projected on sky pixels of the hemispherical images, then the person is exposed to direct sunlight at that time. If the sun position is on non-sky pixels, the person at that time is shaded from sunlight.

The intensity of the sunlight at different times of 1 day varies with the sun elevation angle. In this study, we calculated the weighted sunlight exposure ($E_{w,t}$) at time t as

$$E_{w,t} = B_t \cos\theta_t \tag{1}$$

where θ_t is the sun elevation angle at time t; B_t is the Boolean variable indicating whether the direct sunlight is obstructed or not at time t; if the sunlight is obstructed at time t, B_t equals 0, or B_t is 1. The variations of solar radiation due to cloudiness and other atmospheric conditions are not considered in this study.

3.3 Routing Algorithm for Minimizing Sunlight Exposure

Based on the weighted sunlight exposure, a person's accumulated sunlight exposure E_a from one location at time t_0 to another location at time t_n can be estimated as

$$E_a = \sum_{t=t_0}^{t_n} E_{w,t} = \sum_{t=t_0}^{t_n} B_t \cos\theta_t \tag{2}$$

Figure 4 shows a sequence of hemispherical images with corresponding sun positions overlaid along a street of the study area. The "exposure over distance" parameter α, which indicates the trade-off between sunlight exposure and distance, was used to trade off between distance and sunlight exposure. Therefore, the routing algorithm will find the minimum accumulated sunlight exposure (Min E_a) from several route candidates:

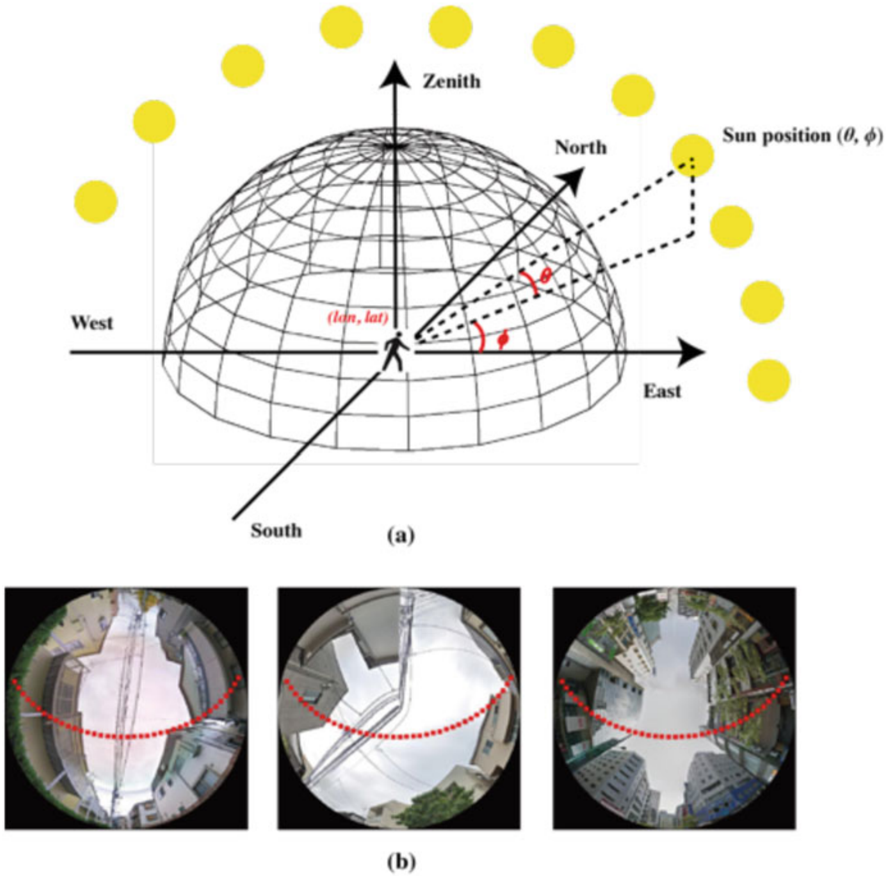


Fig. 3 The geometrical model of the sun (**a**) and the overlay of sun path in 1 day on hemispherical images (**b**)

$$\text{Minimize } E_a = \sum_{t=t_0}^{t_n} \text{dist} \cdot [\alpha B_t \cos\theta_t + (1 - \alpha)] \quad (3)$$

where dist is the distance between two nearby stop points along the route. The trade-off parameter α as 0 indicates the shortest geographical distance path from origin to destination. The trade-off parameter α as 1 indicates route with the minimum sunlight exposure. In this study, we set the trade-off parameter α as 0.5.

Routing for the minimum exposure to sunlight is a time-dependent routing issue, in which the weight of the graph changes over time. Therefore, we first generated the sunlight exposure distribution along streets every 5 min. We assume that the sunlight exposure at each street segment is constant in 5 min. We then applied Dijkstra (1959) algorithm with consideration of the temporal change of the sunlight

Fig. 4 Human exposure to sunlight along one route

exposure for different street segments to find the route with minimum accumulated sunlight exposure for any origin and destination in the study area.

4 Results

Based on the metadata of all available GSV panoramas, we only selected the most recent images captured in leaf-on seasons (April, May, June, July, August, September, and October) in this study. Finally, we collected 45,085 GSV panoramas along streets in the study area. Figure 5 shows the spatial distribution of the finally collected GSV panoramas and the time stamps of those downloaded GSV panoramas. Generally, the GSV panoramas cover most streets in the study area. There is a small region in the northern part of the study area, which is the Meiji Shrine, which has no GSV image coverage. Considering the relatively isolated location of the shrine, we think that the unavailability of GSV images in the small portion of the study area will have not much influence on the results of the routing algorithm in the study area.

Based on the segmented hemispherical images and predicted sun positions over time in the study area, we generated the spatiotemporal distribution of the sunlight exposure at the point level for every 5 min. We then aggregated the point-level sunlight exposure map to street segments by assigning each point to the closest street segments. The sunlight exposure attribute for each street segment is the mean

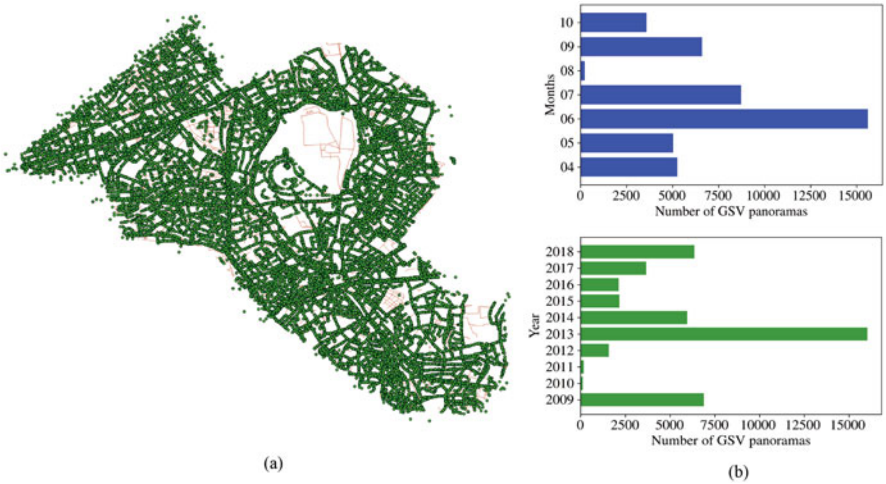


Fig. 5 The spatial distribution of the downloaded Google Street View (GSV) panoramas (**a**) and time stamps (**b**) of those downloaded GSV panoramas

value of all assigned points. Figure 6 shows the spatial distribution of street-level sunlight exposure over time on July 15th, 2018, at 9:00 am, 12:00 pm, 2:00 pm, and 5:00 pm. It can be seen clearly that at 9:00 am and 5:00 pm, those East-West orient streets are exposed to sunlight directly, and streets with other orientations are not directly exposed to sunlight. This is because at sunrise morning and sunset afternoon the sun zenith matches the East-West streets.

In order to compare the results of the shortest geographical distance path and the minimum sunlight exposure path, we randomly created 1000 origin-destination pairs with random starting time from 9 am to 6 pm. Results show that the minimum sunlight exposure paths help to decrease the potential sunlight exposure by 35.23% compared with the shortest geographic distance path. Figure 7 shows 4 of these 1000 paths with the minimum sunlight exposure and trade-off parameter as 0.5 and corresponding paths with the shortest geographic distance from origins to destinations at the starting time of 11 am on July 15th, 2018.

5 Discussion

Too much sunlight exposure would cause human heat stress and unprotected sunlight exposure could cause skin cancers, which is one of the most common cancers. Understanding the spatiotemporal distribution of sunlight exposure would help us to develop strategies to increase human thermal comfort and reduce environmental hazards from the sunlight in cities. Many methods have been developed and applied to protect human beings from too much sunlight exposure, such as planting trees

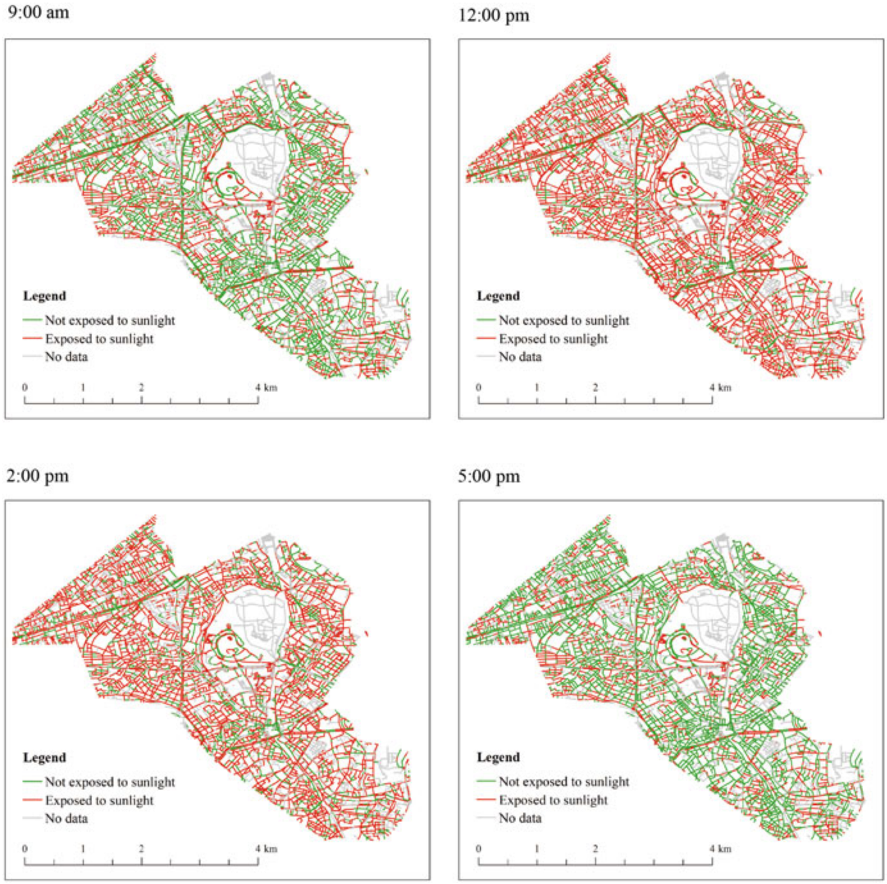


Fig. 6 The spatial distributions of street-level Boolean sunlight exposure variable at 9:00 am, 12:00 pm, 2:00 pm, and 5:00 pm on July 15th, 2018, of the study area

to increase shade, wearing sunscreen, and installing shelters. Different from those plans, this study proposed an individual-level strategy to protect people from too much outdoor sunlight exposure. The Google Street View (GSV) panoramas and deep convolution neural network were used to map the spatiotemporal distribution of the sunlight exposure with consideration of the sun positions and obstruction effects of buildings and tree canopies.

By analyzing the sunlight exposure of 1000 pairs of randomly created origins and destinations in the study area, results show that the proposed routing algorithm can help to reduce the potential sunlight exposure by 35.23% on July 15th, 2018, compared with the shortest geographical distance path. Although the exact number of sunlight exposure reduction would vary for different days and different daily routes of people, this study shows the possibility to significantly reduce the sunlight exposure using the proposed method. This study provides a new way at the

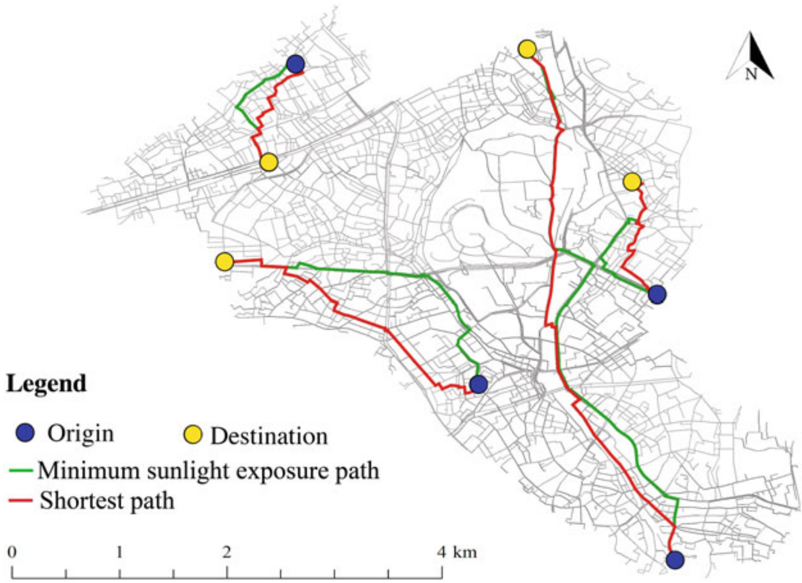


Fig. 7 The paths of minimum sunlight exposure (green line) and the path of shortest geographical distance (red line) of four origin-destination pairs at the randomly starting time from 9 am to 6 pm on July 15th, 2018

individual level to deal with the risks of too much sunlight exposure. Since GSV images are globally available, therefore, the proposed method can be reproduced for other areas to deal with the negative effects of hazards from sunlight. The proposed method provides a bottom-up solution to deal with the human heat stress and unprotected UV exposure caused by outdoor sunlight exposure, because the proposed method would let individuals have the quantitative information about their daily sunlight exposure. With such information available, individuals can change their daily schedule or choose the best route to protect themselves from too much sunlight exposure. This study would also provide an important reference for urban planners and city governments to reduce human sunlight exposure by street designs, such as providing shades and increasing street tree canopies.

The proposed method can generate more accurate sunlight exposure information compared with other digital building height model-based methods since street-level images were used to represent the streetscapes. Using the street-level image-based method is more reasonable to consider the actual human sunlight exposure in street canyons because all kinds of sunlight obstructions are considered using the street-level image-based method.

This study still has some limitations on modeling human sunlight exposure. The street-level image-based method may not be able to perfectly represent pedestrian's sunlight exposure since GSV images were collected in the central part of streets, not sidewalks. Considering the fact that the streets in the study area are very narrow, therefore, it is quite reasonable to use the street-level images to represent pedestrian's exposure to the sunlight. However, the method should be adjusted for other study areas with wide streets and separate sidewalks.

In addition, human thermal stress and UV exposure are influenced by not only the sunlight exposure but also other personal characteristics and meteorological parameters, such as ages, skin types, cloud condition, humidity, wind speed, and air temperature. In this study, only the direct sunlight exposure is considered. The sunlight exposure in street canyons varies spatially and temporally much more than other factors. For one specific site if we assume that other meteorological parameters are constant, therefore, the exposure to the sunlight would be considered as the most important contributing factor of the human thermal comfort and UV exposure. Therefore, it is reasonable to use the sunlight exposure as a surrogate to develop routing algorithm to maximize the thermal comfort and minimize UV exposure for a pedestrian walking from one place to another place. Future studies should also consider more personal characteristics in the sunlight exposure model.

This study used the randomly created sample sites to evaluate the performance of the proposed routing algorithm for reducing the sunlight exposure. In order to better understand the performance of the routing algorithm on local people, future study may also need to use the human actual trajectories. In addition, sunlight exposure is not always a hazard; future studies should also focus on quantitatively managing the sunlight exposure.

6 Conclusions

This study proposed an individual-level and short-term effective strategy to help people get rid of too much sunlight exposure. A time-dependent routing algorithm was developed to minimize human outdoor sunlight exposure based on a bottom-up approach to estimate the spatiotemporal distribution of sunlight exposure along streets using Google Street View. A deep learning-based image segmentation algorithm was used to derive streetscape environment and model the human sunlight exposure within street canyons. Considering the global availability of street-level images around the world, the proposed method would provide us an effective method in dealing with the sunlight exposure protection.

References

Algeciras, J. A. R., Consuegra, L. G., & Matzarakis, A. (2016). Spatial-temporal study on the effects of urban street configurations on human thermal comfort in the world heritage city of Camagüey-Cuba. *Building and Environment*, *101*, 85–101.

Ali-Toudert, F., & Mayer, H. (2006). Numerical study on the effects of aspect ratio and orientation of an urban street canyon on outdoor thermal comfort in hot and dry climate. *Building and Environment*, 41(2), 94–108.

Armstrong, B. K., & Kricker, A. (2001). The epidemiology of UV induced skin cancer. *Journal of Photochemistry and Photobiology B: Biology*, 63(1–3), 8–18.

Brash, D. E., Rudolph, J. A., Simon, J. A., Lin, A., McKenna, G. J., Baden, H. P., ... & Ponten, J. (1991). A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proceedings of the National Academy of Sciences*, 88(22), 10124–10128.

Carrasco-Hernandez, R., Smedley, A. R., & Webb, A. R. (2015). Using urban canyon geometries obtained from Google Street View for atmospheric studies: Potential applications in the calculation of street level total shortwave irradiances. Energy and Buildings, 86, 340–348.

Easter, M. J., & Spies, T. A. (1994). Using hemispherical photography for estimating photosynthetic photon flux density under canopies and in gaps in Douglas-fir forests of the Pacific Northwest. *Canadian Journal of Forest Research*, 24(10), 2050–2058.

Gaffen DJ, Ross R. 1998. Increased summertime heat stress in the United States. *Nature* 396:529–530.

Gong, F. Y., Zeng, Z. C., Zhang, F., Li, X., Ng, E., & Norford, L. K. (2018). Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Building and Environment*, 134, 155–167.

Hodder, S. G., & Parsons, K. (2007). The effects of solar radiation on thermal comfort. *International Journal of Biometeorology*, 51(3), 233–250.

Hwang, R. L., Lin, T. P., & Matzarakis, A. (2011). Seasonal effects of urban street shading on long-term outdoor thermal comfort. *Building and Environment*, 46(4), 863–870.

Johansson, E. (2006). Influence of urban geometry on outdoor thermal comfort in a hot dry climate: a study in Fez, Morocco. *Building and Environment*, 41(10), 1326–1338.

Kurazumi, Y., Kondo, E., Ishii, J., Sakoi, T., Fukagawa, K., Bolashikov, Z.D, et al. (2013), Effect of the environmental stimuli upon the human body in winter outdoor thermal environment, Journal of Environmental and Public Health, 418742, https://doi.org/10.1155/2013/418742, 2013.

Li, X., Ratti, C., & Seiferling, I. (2017, July). Mapping Urban Landscapes Along Streets Using Google Street View. In International Cartographic Conference (pp. 341–356). Springer, Cham.

Li, X., Ratti, C., & Seiferling, I. (2018). Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landscape and Urban Planning*, 169, 81–91.

Lin, T. P., Matzarakis, A., & Hwang, R. L. (2010). Shading effect on long-term outdoor thermal comfort. *Building and Environment*, 45(1), 213–221.

Matzarakis, A., Rutz, F., & Mayer, H. (2007). Modelling radiation fluxes in simple and complex environments—application of the RayMan model. *International Journal of Biometeorology*, 51(4), 323–334.

Matzarakis, A., Rutz, F., & Mayer, H. (2010). Modelling radiation fluxes in simple and complex environments: basics of the RayMan model. *International Journal of Biometeorology*, 54(2), 131–139.

National Oceanic and Atmospheric Administration. 2009. U.S. Heat Stress Index. Heat Stress Datasets and Documentation. Available: http://lwf.ncdc.noaa.gov/temp-and-precip/heat-stress.html#data [accessed 20 February 2009].

Rich, P. M. (1990). Characterizing plant canopies with hemispherical photographs. *Remote Sensing Reviews*, 5(1), 13–29.

Richards, D. R., & Edwards, P. J. (2017). Quantifying street tree regulating ecosystem services using Google Street View. *Ecological Indicators*, 77, 31–40.

Sanusi, R., Johnstone, D., May, P., & Livesley, S. J. (2016). Street orientation and side of the street greatly influence the microclimatic benefits street trees can provide in summer. *Journal of Environmental Quality*, 45(1), 167–174.

Stone, B., Hess, J. J., & Frumkin, H. (2010). Urban form and extreme heat events: are sprawling cities more vulnerable to climate change than compact cities? *Environmental health Perspectives*, 118(10), 1425–1428.

Young, C. (2009). Solar ultraviolet radiation and skin cancer. *Occupational Medicine*, 59(2), 82–88.

Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. “Pyramid scene parsing network.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890. 2017.

Planning and Management of Charging Facilities for Electric Vehicle Sharing

Long He, Guangrui Ma, Wei Qi, Xin Wang, and Shuaikun Hou

1 Introduction

An electrified and shared mobility system is one of the building blocks for a smart-city future, which provides flexible on-demand transit services using a pool of electric vehicles (EVs). However, a smart EV sharing system cannot be built easily. A recent setback occurred in San Diego, where car2go ceased the operations of its EV sharing fleet in 2016. It is caused by insufficient charging which leads to low battery energy levels and thus limited availability of EVs to fulfill travel demands, which can ultimately throttle the operations. To improve the system, EV sharing operators face major challenges. At the strategic level, infrastructure for

L. He
NUS Business School, National University of Singapore, Singapore, Singapore
e-mail: longhe@nus.edu.sg

G. Ma
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
e-mail: guangrui.ma@sjtu.edu.cn

W. Qi (✉)
Desautels Faculty of Management, McGill University, Montreal, QC, Canada
e-mail: wei.qi@mcgill.ca

X. Wang
Department of Industrial and Systems Engineering and Grainger Institute for Engineering, University of Wisconsin-Madison, Madison, WI, USA
e-mail: xin.wang@wisc.edu

S. Hou
Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA
e-mail: shou34@wisc.edu

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_8

battery charging is often scarce. Because of the high construction cost of charging infrastructure and its scarcity, it is essential to carefully decide the locations and quantities of chargers to install. At the operational level, the EV sharing operator has to conduct intertwined repositioning and battery charging operations to meet travel demand. Unbalanced customer trips often lead to either too few or excessive EVs at some locations. It is thus important for the operator to reposition EVs to improve vehicle utilization by meeting more demands at more locations. However, the repositioning of EVs is much more challenging because EVs take significantly longer charge time, yet with shorter per-charge range. The energy remaining in the battery, hereafter referred to as the *energy level*, is critical to demand fulfillment and to the feasibility of repositioning. A salient feature of EV sharing is that the energy levels of vehicle flows should be closely tracked to ensure charging EVs timely, as well as to improve demand fulfillment.

We also reviewed recent literature on the topics of infrastructure planning for electrified car sharing systems and EV service planning and operations. The majority of literature on vehicle sharing service systems has limitations on their work including not considering battery charging operations and other problems. Thus, we briefly summarize some developments in EV charging area. There are few related studies in this area. Boyacı et al. (2015) develop a planning model to analyze station-based EV sharing, which takes station location, parking space allocation, fleet size management, and vehicle relocation into consideration. He et al. (2017) build a mathematical programming model to solve the optimal service region planning for EV sharing. Their model considered customer subscription decisions and fleet operations, while not considering heterogeneous energy levels of EVs and simply assume that 20% of all EVs arriving at each location enter charging process and all EVs available for customers are fully charged. Moreover, they consider demand fulfillment as a centralized decision by the operator, which is not realistic. Abouee Mehrizi et al. (2018) propose an analytical model to discuss the viability of using EVs in a car sharing system by considering an aggregated market without heterogeneities across locations and EVs. Mak et al. (2013) optimize battery swapping locations along highways to accommodate exogenous EV flows. However, it requires negligible time. Lim et al. (2015) compare potential mass EV adoption between selling and leasing business settings with consumer anxieties about the range and resale of EVs. Jiang and Powell (2016) propose to solve a risk-averse Markov decision process model for dynamically charging EVs in the presence of electricity spot price variations. Zhang et al. (2017) address the coupling between transportation and electrical grid networks when planning EV charging facilities. Charging requirement also brings new challenges to freight logistics services when employing EVs. Schneider et al. (2014) and Desaulniers et al. (2016) aim at solving EV routing problems with charging schemes under time window constraints and they propose optimization algorithms.

In this chapter, based on the work of He et al. (2021), we describe models, analytics, and insights to design a better EV sharing system. The objective is to maximize the revenue net of the infrastructure investment and operating cost while satisfying urban mobility demands in a shared and electrified system. We present an integrated model to determine the locations and sizes of battery charging sites, along with the coupled fleet charging and repositioning operations. Also, a case study based on real data is conducted to understand if car2go could have performed better. More importantly, we provide managerial insights into the future development of EV sharing for cities and companies that pursue this mobility paradigm.

2 EV Sharing System Planning and Operations

The fleet operations in the EV sharing system are modeled as a queueing network. Planning charging infrastructure contains charging site selection and charger installation to form the network. He et al. (2021) use binary decision variable X_i to denote whether the operator deploys any charger in zone $i \in \mathscr{I}$, where $\mathscr{I}$ is the set of zones that form the service region. At selected site i, i.e., $X_i = 1$, the operator denotes the number of chargers by Y_i. The demand zones are considered as a multi-class open queueing network consisting of single-server nodes, and infinite-server nodes for the roads, and of multi-server nodes for the charging sites. EVs are viewed as the entities in the queueing network. Each type of node and their connections are introduced below before the detailed discussions in the following subsections.

- Each demand zone is considered as a single "server". An "arrival" event indicates that an EV is dropped off in a zone and a "departure" event indicates that an EV leaves a zone. The available EVs that wait in zone i will be picked up by customers at the service rate that is endogenously determined by the demand rate and customers' EV picking behavior. Note that EVs are differentiated by their energy levels and customers tend to pick available EVs with the highest energy. Thus, each demand node is modeled as a multi-class $M/M/1$ queue with preemptive-repeat priority and class-dependent service rates in Sect. 2.1.
- The roads are modeled as $M/G/\infty$ nodes with generally distributed travel times, as EVs leaving a demand zone can immediately get on the road (enter "service").
- The charging node in zone i, if available, is represented by an $M/G/Y_i$ queue where Y_i specifies the number of chargers. Its service time distribution depends on the specific charging policy, which we will analyze in Sect. 2.2 under a charging up-to policy motivated by the data.
- Nodes in the network interact with each other through the movements of vehicles.
- The vehicle arrivals and departures are modeled as a Poisson distribution. We examine it using an operational data set from car2go in San Diego.

2.1 *Customers' EV Picking*

It is important for an EV sharing operator to understand how customers choose EVs to fulfill their travel needs. In this chapter, EVs are differentiated by their energy

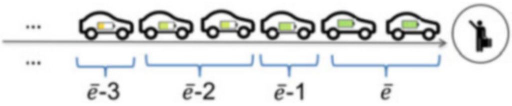


Fig. 1 EV picking by "risk-averse" customers with range anxiety. Adapted from 'Charging Electric Vehicle Sharing Fleet', by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 11

levels, which is indicated by e. The range of energy levels is divided into a finite set $\mathscr{E} = \{0, 1, \ldots, \bar{e}\}$, where $\bar{e}$ is the highest energy level. Through the website or mobile applications, customers are able to access to the energy level information of EVs in the EV sharing system. Therefore, He et al. (2021) use e_{ij} to denote the energy consumed by a trip from i to j. A customer who is traveling from zone i to zone j will seek an available EV in zone i that has at least e_{ij} amount of energy remaining. If there is no available EV or all available EVs do not have sufficient energy (i.e., $e < e_{ij}$) in zone i, the demand will be lost. Customers' EV picking behavior is modeled based on the risk-averse mode, which assumes that the customer will choose the EV with the highest energy level to minimize the risk of battery depletion before the trip ends. The risk-averse mode is empirically verified in Kim et al. (2018).

The demand fulfillment process becomes highest-energy-first-out (HEFO) as shown in Fig. 1 with the risk-averse mode of EV picking, which is modeled as a multi-class M/M/1 queue with preemptive-repeat priority and class-dependent service rates. The service rate for an EV with energy level e is $\hat{d}_{ie} = \sum_{j \in \mathscr{J}_{ie}} d_{ij}$, where d_{ij} is the demand rate for trips from zone i to zone j. $\mathscr{J}_{ie} = \left\{j : e_{ij} \leq e\right\}$ is the set of destinations reachable by energy level e from origin i. All EVs are divided into $\bar{e}$ classes according to their energy levels, i.e., class e for EVs with energy level e. Under the risk-averse mode, an EV can be "served by zone i" only when all EVs of higher class (i.e., with higher energy levels) have been taken, which clearly suggests the preemptive priority rule—class $\bar{e}$ has the highest priority and class 0 has the lowest priority.

Let ψ_{ije} be the fulfilled demand rate for customers traveling from i to j using EVs with energy level e. The load of class e can be defined as

$$\alpha_{ie} = \frac{\sum_{j \in \mathscr{J}_{ie}} \psi_{ije}}{\hat{d}_{ie}} \tag{1}$$

for zone i. Also, the operator is not able to ration the demand to fulfill by trip destinations because customers are not required to inform the operator about their destinations. In this way, the satisfied customer trips follow the proportional rule as

$$\psi_{ije} = \frac{d_{ij}}{\hat{d}_{ie}} \sum_{j \in \mathscr{J}_{ie}} \psi_{ije}. \tag{2}$$

Next, the expected queue lengths for EVs with different energy levels in zone i is denoted by load α_{ie}. The queue length formula for multi-class M/G/1 queue with preemptive priority in Shortle et al. (2018) is adopted. With the mean $\frac{1}{\hat{d}_{ie}}$ and the second moment $\frac{2}{\hat{d}_{ie}^2}$ of the exponential service time of class e, the average number (queue length) of EVs with energy level e in zone i is given by

$$L_{ie} = \frac{\alpha_{ie}\left[1 - \sum_{e'=e}^{\bar{e}} \alpha_{ie'} + \hat{d}_{ie} \sum_{e'=e}^{\bar{e}} \frac{\alpha_{ie'}}{\hat{d}_{ie'}}\right]}{\left(1 - \alpha_{i\bar{e}} - \alpha_{i(\bar{e}-1)} - \cdots - \alpha_{i(e+1)}\right)\left(1 - \alpha_{i\bar{e}} - \alpha_{i(\bar{e}-1)} - \cdots - \alpha_{ie}\right)}, \ \forall e < \bar{e}, \tag{3}$$

$$L_{i\bar{e}} = \frac{\alpha_{i\bar{e}}}{1 - \alpha_{i\bar{e}}}. \tag{4}$$

2.2 Charging Process

He et al. (2021) analyzed the energy levels of EVs of 1472 car2go trips and decided to use a *proactive* charging policy where the operator can choose to charge EVs with different energy levels (lower than $\bar{e}$). Also, they adopt a *charge up-to policy* to reflect when to stop charging, which means that the operator charges all EVs to energy level $\bar{e}$. The parameter $\bar{e}$ indicates the highest energy level. In practice, $\bar{e}$ can potentially vary by time or location (which a direct extension of our model can capture).

The charging process at each site i with Y_i chargers is modeled as an $M/G/Y_i$ queue. Adopting proactive charge up-to policy, the EVs that arrive with energy level e is a Poisson distribution with rate λ_{ie} (which is controlled by the operator). The charging time is $\frac{\bar{e}-e}{u}$, where u denotes the charging power. Therefore, the charging time of any EV follows a discrete distribution, being $\frac{\bar{e}-e}{u}$ with probability $\frac{\lambda_{ie}}{\sum_{e<\bar{e}} \lambda_{ie}}$, with mean $a_i = \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \frac{\lambda_{ie}}{\sum_{e<\bar{e}} \lambda_{ie}}$ and variance $b_i^2 = \sum_{e<\bar{e}} \frac{(\bar{e}-e)^2}{u^2} \frac{\lambda_{ie}}{\sum_{e<\bar{e}} \lambda_{ie}} - a_i^2$. Then, the expected waiting time in the queue (excluding the charging time) is approximated by W_i^q, using the heavy-traffic formula from Whitt (1993) as follows:

$$\mathbb{E}[W_i^q] = \frac{\rho_i a_i}{Y_i(1-\rho_i)} \frac{c_a^2 + c_s^2}{2},$$

where the utilization ratio $\rho_i = \frac{a_i \sum_{e<\bar{e}} \lambda_{ie}}{Y_i}$, $c_s^2 = \frac{b_i^2}{a_i^2}$ and $c_a^2 = 1$. The heavy-traffic approximation is asymptotically accurate as ρ_i approaches 1.

Therefore, the expected time an EV spends at site i, including both the time in the queue and being charged, is given by

$$\mathbb{E}[W_i] = \mathbb{E}[W_i^q] + a_i = \frac{\sum_{e<\bar{e}} \frac{(\bar{e}-e)^2}{u^2} \lambda_{ie}}{2Y_i^2 - 2Y_i \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie}} + \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \frac{\lambda_{ie}}{\sum_{e<\bar{e}} \lambda_{ie}}.$$

By Little's law, the expected number of vehicles Q_i in charging site i is given by

$$Q_i = (\sum_{e<\bar{e}} \lambda_{ie}) \mathbb{E}[W_i] = \frac{(\sum_{e<\bar{e}} \lambda_{ie}) \sum_{e<\bar{e}} \frac{(\bar{e}-e)^2}{u^2} \lambda_{ie}}{2Y_i^2 - 2Y_i \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie}} + \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie}. \tag{5}$$

Since the utilization ratio $\rho_i \leq 1$, we have

$$Y_i \geq \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie}. \tag{6}$$

2.3 *The Optimization Model*

He et al. (2021) develop an integrated optimization model for design and operations of a system with given m EVs. The EV sharing fleet moves across and within zones. EV movements consist of both transit trips by the customers and repositioning trips by the operator, which are aggregated EV flows at different energy levels. Because only EVs with energy level $e \geq e_{ij}$ are feasible for such a trip, we have $\psi_{ije} = 0$ for $e < e_{ij}$. In order to rebalance the system if customers transit trips lead to asymmetric EV flows, the operator needs to reposition EVs from zone i to j with rate ϕ_{ije} for EVs with energy level e. Suppose an EV consumes e^r_{ij} energy for repositioning from zone i to j, $\phi_{ije} = 0$ for $e < e^r_{ij}$. The operator also dispatches workers to move EVs within each zone: EVs with energy level $e < \bar{e}$ enter the charging site in zone i (if opened) with rate λ_{ie} and depart from the charging process with rate η_i after reaching energy level $\bar{e}$.

The following *energy-flow balance equations* characterize the EV flows in each zone and in the charging sites:

$$\sum_{j\in\mathscr{I}} \left(\psi_{ji(e+e_{ji})} + \phi_{ji(e+e^r_{ji})}\right) = \sum_{j\in\mathscr{I}} \left(\psi_{ije} + \phi_{ije}\right) + \lambda_{ie}, \quad \forall i \in \mathscr{I}, e \in \mathscr{E}\backslash\bar{e} \tag{7}$$

$$\eta_i = \sum_{j\in\mathscr{I}} \left(\psi_{ij\bar{e}} + \phi_{ij\bar{e}}\right) + \lambda_{i\bar{e}}, \quad \forall i \in \mathscr{I} \tag{8}$$

$$\eta_i = \sum_{e\in\mathscr{E}} \lambda_{ie}, \quad \forall i \in \mathscr{I}. \tag{9}$$

The balance equations (7) and (8) indicate the EV flows with energy levels $e < \bar{e}$ and $\bar{e}$, respectively, for each zone i. In Eq. (7), the term $\sum_{j\in\mathscr{I}}\left(\psi_{ji(e+e_{ji})}+\phi_{ji(e+e^r_{ji})}\right)$ summarizes the total EV inflows into zone i, whose energy level on arrival is e. Right-hand side (RHS) $\sum_{j\in\mathscr{I}}\left(\psi_{ije}+\phi_{ije}\right)+\lambda_{ie}$ denotes the total outflow from zone i by transit trips, repositioning trips for fleet rebalance and to the charging site at zone i. Note that this equation concerns about EV inflows from other zones and therefore their energy levels are less than $\bar{e}$. For EVs with energy level $\bar{e}$, Eq. (8) indicates that the inflow comes only from the charging site in zone i with rate η_i and the term on the RHS $\sum_{j\in\mathscr{I}}\left(\psi_{ij\bar{e}}+\phi_{ij\bar{e}}\right)+\lambda_{i\bar{e}}$ is the total EV outflow with energy level $\bar{e}$. The last Eq. (9) shows the flow balance at the charging site i.

Then, the customers' EV picking and the operator repositioning process are integrated using a fleet size constraint. He et al. (2021) use Little's law to derive the expected numbers of EVs traveling from zone i to j with initial energy level e are $t_{ij}\psi_{ije}$ for customer transit and $s_{ij}\phi_{ije}$ for repositioning, where t_{ij} and s_{ij} are the expected travel time in transit and repositioning between zones, respectively. Let s_{ii} be the expected travel time for EVs entering and leaving the charging sites. The expected number of EVs repositioned to and from the charging sites is given by $s_{ii}\left(\sum_{e<\bar{e}}\lambda_{ie}+\eta_i\right)$ in zone i. The sum of all the above EV quantities is no larger than the fleet size m, which the following fleet size constraint imposes:

$$\sum_{i\in\mathscr{I}}\left(Q_i+\sum_{e\in\mathscr{E}}L_{ie}\right)+\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{I}}\sum_{e\in\mathscr{E}}\left(t_{ij}\psi_{ije}+s_{ij}\phi_{ije}\right)$$
$$+\sum_{i\in\mathscr{I}}s_{ii}\left(\sum_{e<\bar{e}}\lambda_{ie}+\eta_i\right)\le m. \tag{10}$$

Let p denotes the revenue from customer usage per unit time per EV, c denotes the repositioning cost per unit time per EV, f_i denotes the fixed setup cost of a charging site in zone i and g the installation cost per charger. Suppose the maximum number of chargers allowed is $\bar{y}_i$ in zone i. The charging infrastructure planning and fleet operations problem for EV sharing is formulated as a nonlinear optimization problem (NLP):

$$\max_{\substack{X,Y,\phi,\psi\\ \alpha,\lambda,\eta,L,Q}} \quad p\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{I}}\sum_{e\in\mathscr{E}}t_{ij}\psi_{ije}-c\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{I}}\sum_{e\in\mathscr{E}}s_{ij}\phi_{ije}$$
$$-c\sum_{i\in\mathscr{I}}s_{ii}\left(\sum_{e<\bar{e}}\lambda_{ie}+\eta_i\right)-\sum_{i\in\mathscr{I}}\left(f_iX_i+gY_i\right)$$
$$\text{s.t.}\quad \text{Constraints}(1)-(10),$$
$$Y_i\le\bar{y}_iX_i,\quad \forall i\in\mathscr{I}$$

$$\phi_{ije} = 0, \quad \forall i, j \in \mathscr{I}, e < e^r_{ij} \quad or \quad e > \bar{e}$$

$$\psi_{ije} = 0, \quad \forall i, j \in \mathscr{I}, e < e_{ij} \quad or \quad e > \bar{e}$$

$$0 \le \alpha_{ie} \le 1, \forall i \in \mathscr{I}, e \in \mathscr{E}$$

$$Y_i, \phi_{ije}, \psi_{ije}, \lambda_{ie}, \eta_i, L_{ie}, Q_i \ge 0, \quad \forall i, j \in \mathscr{I}, e \in \mathscr{E}$$

$$X_i \in \{0, 1\}, \quad \forall i \in \mathscr{I}.$$

Here, the objective is to maximize the operator's expected annual profit, which is the revenue from transit trips by customers, the repositioning cost by workers and the investment in charging infrastructure. Vehicle flow rates ($\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\eta}$) and fixed costs ($f, g$) are all annualized. The additional constraint $Y_i \le \bar{y}_i X_i$ ensures that no charger will be installed without a charging site. Pointwise stationary approximation approach (Green and Kolesar 1991) is adopted to incorporate system dynamics over multiple periods.

2.4 Solution Approach

The problem (NLP) is nonlinear and nonconvex because of the fractional constraints (3)–(5) with quadratic terms that define queue lengths L_{ie} and Q_i. Thus, He et al. (2021) propose alternative formulations by overestimating (underestimating, resp.) L_{ie} and Q_e to generate lower (upper, resp.) bound of the optimal (NLP) objective value.

2.4.1 Lower Bound

First, He et al. (2021) develop a transformation of L_{ie} into $\hat{L}_{ie}$ in Lemma 1. Lemma 1 shows that $\sum_{e \in \mathscr{E}} L_{ie}$ can be effectively simplified as a sum of linear ratios.

Lemma 1 *The average number of idle EVs waiting in zone i is given by*

$$\sum_{e \in \mathscr{E}} L_{ie} = \sum_{e \in \mathscr{E}} \hat{L}_{ie}, \ \textit{where} \ \hat{L}_{ie} = \begin{cases} 0, & \text{if} \quad e = 0, \\ \frac{(\hat{d}_{ie} - \hat{d}_{i(e-1)}) \sum_{e'=e}^{\bar{e}} \frac{\alpha_{ie'}}{\hat{d}_{ie'}}}{1 - \sum_{e'=e}^{\bar{e}} \alpha_{ie'}}, & \text{if} \quad e = 1, \ldots, \bar{e}. \end{cases} \tag{11}$$

Rewrite the definition of $\hat{L}_{ie}$ as

$$\hat{L}_{ie} \geq \begin{cases} 0, & \text{if} \quad e = 0, \\ \frac{(\hat{d}_{ie} - \hat{d}_{i(e-1)}) \sum_{e'=e}^{\bar{e}} \frac{\alpha_{ie'}}{\hat{d}_{ie'}}}{1 - \sum_{e'=e}^{\bar{e}} \alpha_{ie'}}, & \text{if} \quad e = 1, \ldots, \bar{e}. \end{cases} \tag{12}$$

The overestimation of $\hat{L}_{ie}$ is built in Lemma 2 and those of Q_i in Lemma 3 as follows.

Lemma 2 *For any $e > 0$ and any constant $\hat{m}_{ie} > 0$, if $\hat{L}_{ie}$ satisfies the constraint*

$$\left(\hat{L}_{ie} + (1 - \sum_{e'=e}^{\bar{e}} \alpha_{ie'})\right)^2 \geq \frac{\left((\hat{d}_{ie} - \hat{d}_{i(e-1)}) \sum_{e'=e}^{\bar{e}} \frac{\alpha_{ie'}}{\hat{d}_{ie'}} + \hat{m}_{ie}\right)^2}{\hat{m}_{ie}} + \left(\hat{L}_{ie} - (1 - \sum_{e'=e}^{\bar{e}} \alpha_{ie'})\right)^2, \tag{13}$$

then $\hat{L}_{ie}$ also satisfies the constraint (12).

Lemma 3 *For any constants $\hat{n}_i, \hat{w}_i > 0$, if Q_i satisfies the constraints*

$$\begin{cases} \left(Q_i - \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie} + Z_i\right)^2 \geq \left(\frac{\sum_{e<\bar{e}} \frac{(\bar{e}-e)^2}{\hat{n}_i} \lambda_{ie} + \hat{n}_i \sum_{e<\bar{e}} \lambda_{ie}}{u}\right)^2 \\ \qquad\qquad + \left(Q_i - \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie} - Z_i\right)^2, \\ \left(2Y_i - \sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie}\right)^2 \geq \frac{(Z_i + X_i \hat{w}_i)^2}{2\hat{w}_i} + \left(\sum_{e<\bar{e}} \frac{\bar{e}-e}{u} \lambda_{ie}\right)^2, \end{cases} \tag{14}$$

then Q_i also satisfies the modified constraint (5) *with a replacement of "=" with "≥".*

Therefore, a lower bound on the optimal value of NLP is found by using the constraints from Lemmas 2 and 3.

Proposition 1 *A lower-bound problem for (NLP) can be formulated as a mixed-integer second order cone program (MISOCP), by replacing $\sum_{e \in \mathscr{E}} L_{ie}$ with $\sum_{e \in \mathscr{E}} \hat{L}_{ie}$ in* (10) *and replacing constraints* (3)–(5) *in (NLP) with constraints* (13) *and* (14).

He et al. (2021) name the resulting formulation (SOCP-LB), which is a MISOCP since (13) and (14) can be converted into SOC constraints in the standard 2-norm form. Also, all the other constraints and the objective function are linear in the decision variables. The formulation is a feasible solution to (NLP) because the optimal objective value of (SOCP-LB) is no greater than that of (NLP).

2.4.2 Upper Bound

He et al. (2021) next underestimated $\sum_{e\in\mathscr{E}} L_{ie}$ and Q_i to obtain an upper bound for (NLP).

Lemma 4 *If L_{ie} satisfies constraints (3) and (4) in (NLP), then $\forall e^c \in \{1, 2, \ldots, \bar{e}\}$, $\sum_{e\in\mathscr{E}} L_{ie}$ satisfies the SOC representable constraint*

$$\left(\frac{\sum_{e\in\mathscr{E}} L_{ie} + 2 - \sum_{e=e^c}^{\bar{e}} \alpha_{ie}}{2}\right)^2 \geq \left(\frac{\sum_{e\in\mathscr{E}} L_{ie} + \sum_{e=e^c}^{\bar{e}} \alpha_{ie}}{2}\right)^2 + \frac{\hat{d}_{e^c}}{\hat{d}_{\bar{e}}}. \tag{15}$$

Next, underestimate Q_i by dropping the fractional term in constraint (5), as follows:

$$Q_i \geq \sum_{e<\bar{e}} \frac{\bar{e} - e}{u} \lambda_{ie}. \tag{16}$$

Therefore, the upper bound for (NLP) is found by using Lemma 4 and constraint (16), which is shown in Proposition 2:

Proposition 2 *The upper-bound problem for (NLP) can be formulated as an MISOCP by replacing constraints (3)–(4) and (5) in (NLP) with constraints (15) and (16).*

He et al. (2021) name the resulting formulation (SOCP-UB). The optimal value of (SOCP-UB) is higher than that of (NLP).

2.5 *Numerical Results*

He et al. (2021) demonstrate their charging planning and operation framework with a case study in San Diego, California, where there is an urban setting of 16 ZIP Code zones. Car2go ran an EV sharing program in 2011–2016 before it quit the San Diego market at the end of 2016. The study identifies 19,380 EV sharing trips and 379 EVs from the data set.

By solving the MISOCP proposed in the previous section, the result suggests that *the operator can serve 95.3% of the total demand by using only three charging sites with proactive repositioning activities.* Figure 2 shows that our proposed design consists of three charging sites in zones 1, 7, and 14, which are either in or next to the zones with high demand flow volume (including both inflows and outflows). Having 40 chargers in zone 1 (downtown San Diego), the operator can utilize the high traffic by customer trips to bring EVs to chargers. Zones 7 and 14 are both recommended to have 16 chargers such that these chargers are easily accessible from nearby zones in the west and east, respectively.

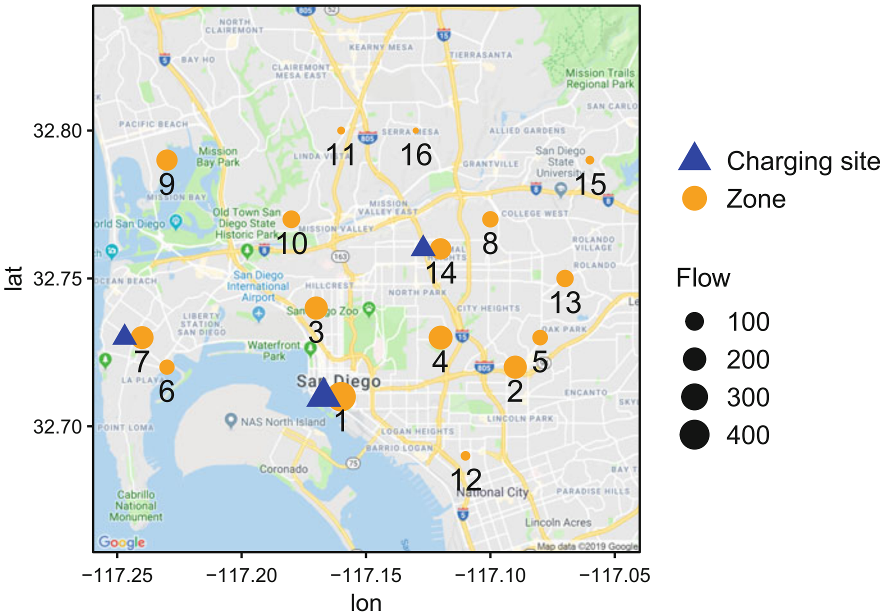


Fig. 2 Charging sites and daily demand flow volume of each zone in San Diego. Adapted from 'Charging Electric Vehicle Sharing Fleet', by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 21

Based on extensive numerical experiment, He et al. (2021) provide several major insights as below.

1. Car2go should ideally proactively charge EVs from varying energy levels contingent upon the operational status of the entire system. Alternatively, the charging activation threshold could be increased to 40%.
2. It is important to ensure sufficient charger availability when collaborating with a public charger network.
3. The EV fleet size could be reduced by up to 20% to achieve optimal profit.

3 Specifications

3.1 Fleet Size

To study the sensitivity of the bound performance related to fleet size, He et al. (2021) solve both the LB and UB profits using (SOCP-LB) and (SOCP-UB) formulations with different fleet size. Figure 3 shows that the optimality gap becomes reasonably small—less than 8% once the fleet size becomes at least 70%

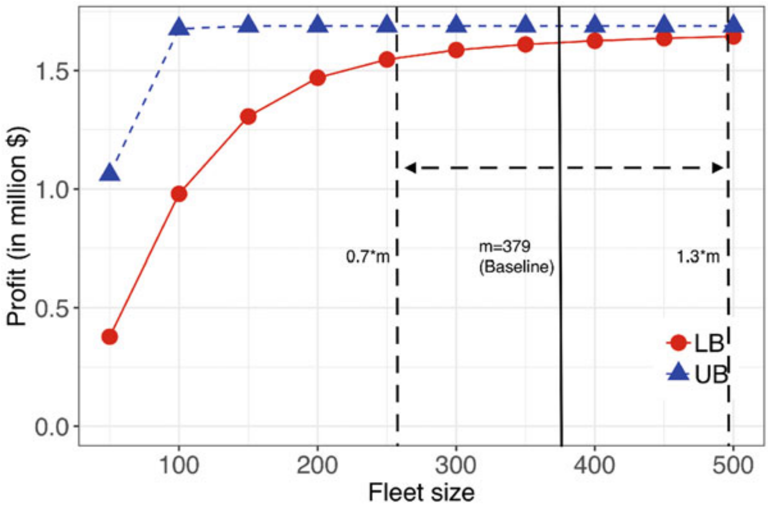


Fig. 3 LB and UB profits in fleet size. Adapted from 'Charging Electric Vehicle Sharing Fleet', by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 22

of the baseline level. Figure 3 also indicates that *car2go may have oversized its EV fleet*. Only 79.2% the baseline size of car2go's fleet, which is a fleet size of 300, is sufficient to generate more than 98.0% of the baseline LB profit (or 93.6% of the baseline UB profit). If more EVs are added, the LB and UB profits are increased by a little. The optimality gap widens when the fleet size drops below 70% of the baseline level, making it more difficult to infer the profitability. However, given the trip demand volume from the data, those scenarios are unrealistic as the fleet sizes deviate too much from the actual baseline level. Other factors that impact car2go's performance are also investigated. However, the optimality gaps between the LB and UB profits are less than 5%, suggesting that the gap is insensitive to the variation of the other parameter values within the considered ranges.

3.2 Proactive vs Threshold-Activated Charging Up-to Policies

He et al. (2021) also investigate whether car2go could have improved its charging operations. The charging operations in the proposed solution are examined by checking the energy levels of EVs. Figure 4a shows the energy level distribution of EVs entering charging sites, that is, $\sum_{i\in\mathscr{I}} \lambda_{ie}$ aggregated over all periods and then normalized into proportions with respect to each energy level e (here the battery capacities are evenly divided into 15 energy levels with the charging up-to level $\bar{e} = 15$). The figure indicates that *it is optimal to proactively charge EVs from varying energy levels*.

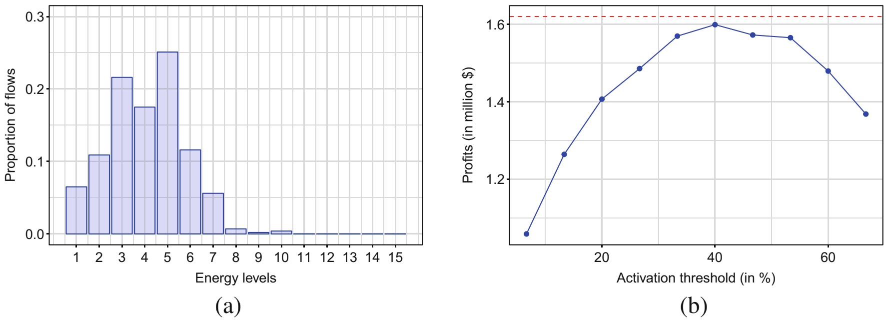


Fig. 4 Comparison between the proactive charging policy and threshold-activated charging policies. (**a**) Energy level distribution of EV flows (aggregated over all periods) entering charging sites under the proactive charging policy. (**b**) Profits with different charging activation thresholds (blue dotted line) and the optimal profit with proactive charging (red dashed line). Adapted from 'Charging Electric Vehicle Sharing Fleet', by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 23

Despite its optimality, the proactive charging policy might be challenging to implement. However, the threshold-activated charging policy is more implementable given its "single-parameter" simplicity. Therefore, we are interested in whether the EV sharing service provider can keep employing a threshold-activated charging policy, while earning a higher profit by simply adjusting the charging activation threshold.

Figure 4b shows that the profit is remarkably sensitive to the threshold, peaking at the 40% level. The operator may gain as much as 98.1% of the optimal profit with the proactive charging policy by raising the charging activation threshold from 20 to 40% (with the charging infrastructure re-optimized). Both the 40%-threshold-activated charging policy and the proactive charging policy outperform the 20%-threshold-activated charging policy. Therefore, *the EV sharing service provider may recover a near-optimal profit with a threshold-activated charging policy by carefully choosing the threshold level*.

3.3 Private vs. Public Chargers: Location and Availability

He et al. (2021) investigate the relationship between public chargers at non-optimized locations and operating profits. A major challenge of using public chargers is that their availability is not guaranteed. To illustrate the impact of charger availability on the trade-off between using public and private chargers, we provide two scenarios: (1) When using the public charging network, the operator has access to $r\%$ of the 386 public chargers at given locations, where $r\%$ is the average charger availability. (2) When using a private charging network, the firm is able to optimize the allocation of equivalent $386 \times r\%$ number of chargers. In each scenario, He

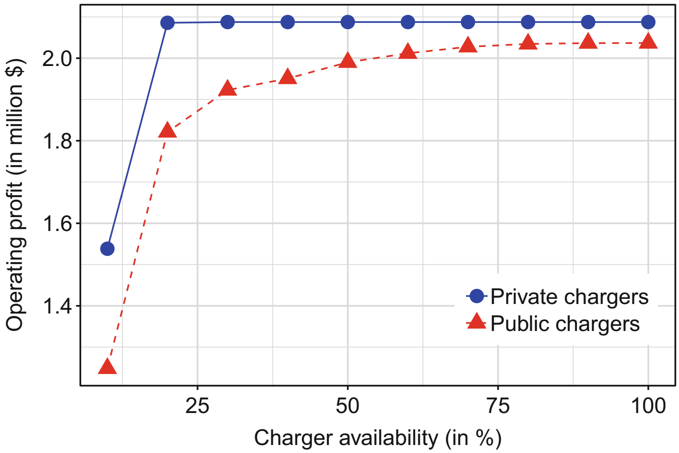


Fig. 5 Operating profits of using public and private chargers. Adapted from 'Charging Electric Vehicle Sharing Fleet', by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 25

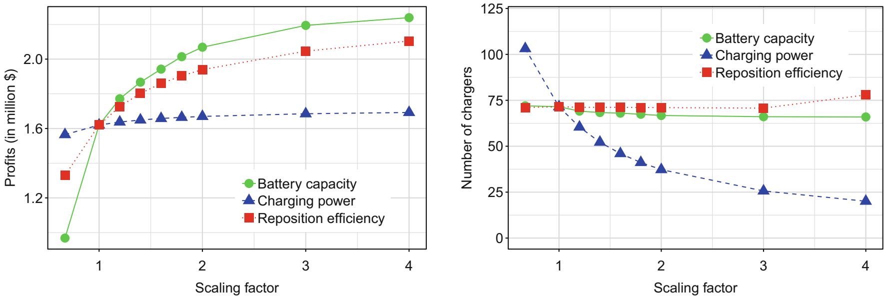
et al. (2021) solve a (SOCP-LB) which eliminates the fixed and variable charger investment costs. Figure 5 shows the operating profits at different levels of charger availability $r\%$ for both cases.

Figure 5 indicates that *when using a public charging network, the charger availability is important to the operating profits.* Car2go should set up its own charging network when the availability of the public charging network is lower than 25%, because the operating profit is much lower than using private chargers. If there are sufficient public chargers, car2go would tend to use public chargers, and effectively reposit EVs at the right locations.

3.4 Technological Advancements

EV sharing operators have gone through significant advancements in EV technologies: First, the charging technology is developing quickly. The rated level-2 charging power can increase to more than 19kW, which is about six times the current 3.3kW for car2go's Smart Electric Drive (Yilmaz and Krein 2013). Second, R&D in the energy density of lithium-ion battery enables longer per-charge range for EVs (Blomgren 2017). Third, combing autonomous vehicle technologies with EVs is a future trend (Fortune 2016). Autonomous EV repositioning and charging are more financially efficient than using human labor.

He et al. (2021) examine the impact of these technological advancements on EV sharing service operations. Figure 6 shows the profit and the total number of chargers in need in response to the parameter variations. The technological factors



(a) Changes in profit (b) Changes in the number of chargers

Fig. 6 Impacts of charging power, battery capacity and repositioning cost efficiency on (**a**) the profit and (**b**) the number of chargers in need. Adapted from ‘Charging Electric Vehicle Sharing Fleet’, by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 26

have impacts as follows: (1) The expected charging time and number of chargers will be reduced if the charging power is significantly increased. However, the profit does not have a significant increase. (2) The profit will be increased if battery capacity and the EV range are increased. However, it has marginal impact on the number of chargers needed. (3) The repositioning cost will be reduced if the per-unit repositioning cost is directly reduced. Thus, charging sites can be built at less expensive locations without a significant increase of repositioning cost, although the repositioning trips become longer.

Therefore, when chargers are limited in the early stage, the EV sharing operator should first consider enhancing the charging power to alleviate its heavy dependence on charging infrastructure. When chargers become increasingly available, the operator should consider expanding battery capacity and reducing the repositioning cost to enhance its profitability.

3.5 Urban Spatial Structure

Finally, He et al. (2021) study how EV sharing service operations are affected by urban spatial structure, especially the centrality of the trip pattern. Let matrix $\mathbf{A}^{\xi}$ denotes the trip pattern of a city with n zones. The element a_{ij}^{ξ} in $\mathbf{A}^{\xi}$ is the probability that this trip will start from zone i and end at zone j. Moreover, $\mathbf{A}^{\xi}$ is parameterized by $\xi \in [0, 1]$:

$$\mathbf{A}^{\xi} = (1 - \xi)\mathbf{A}^{D} + \xi \mathbf{A}^{C},$$

where

$$\mathbf{A}^D = \begin{bmatrix} 0 & 1/(n-1) & \cdots & 1/(n-1) \\ 1/(n-1) & 0 & \cdots & 1/(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ 1/(n-1) & 1/(n-1) & \cdots & 0 \end{bmatrix} \text{ and}$$

$$\mathbf{A}^C = \begin{bmatrix} 0 & 1/(n-1) & \cdots & 1/(n-1) \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}.$$

$\mathbf{A}^\xi$ is a convex combination of $\mathbf{A}^D$ and $\mathbf{A}^C$, where $\mathbf{A}^D$ represents the fully decentralized scenario in which all destinations are equally likely for a one-way trip from any origin, and $\mathbf{A}^C$ represents the fully centralized scenario where zone 1 as the central district is the only destination for all trips from other zones and all trips from zone 1 are equally likely to end at any other zone. Hence, ξ captures urban spatial centrality: As ξ increases from 0 to 1, the trip pattern becomes more centralized around zone 1.

The authors use $\mathbf{A}^\xi$ to generate numerical instances with different degrees of centrality. Let d_i^k be the rate of customer trip demands originating from zone i in period k. They keep d_i^k the same as in our baseline setting, but split d_i^k to each zone j according to $d_{ij}^k = d_i^k a_{ij}^\xi$. Then they solve (SOCP-LB) for six different values of centrality ξ ranging from 0 to 1.

Based on the study, we know that the EV sharing operator may earn more profit as the urban centrality ξ increases. As Fig. 7 shows, this profit increase is

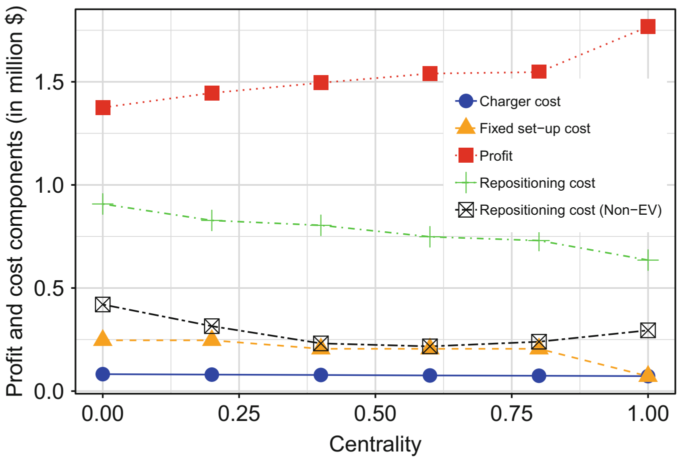
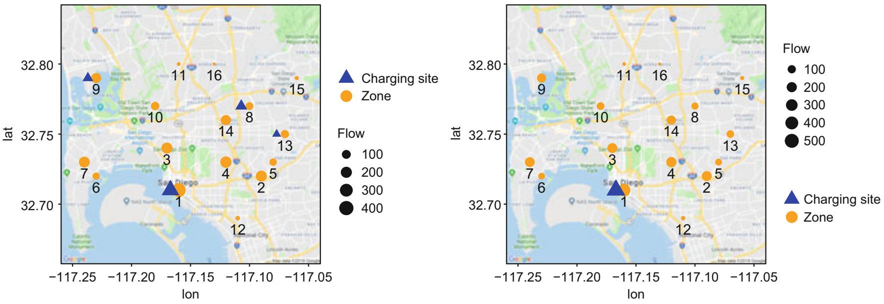


Fig. 7 Profit and cost components in demand pattern measured by centrality ξ. Adapted from 'Charging Electric Vehicle Sharing Fleet', by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 28



(a) (b)

Fig. 8 Layouts of charging sites under different demand patterns measured by centrality ξ. (**a**) 83 chargers on 4 sites for centrality $\xi = 0$. (**b**) 73 chargers on 1 site for centrality $\xi = 1$. Adapted from ‘Charging Electric Vehicle Sharing Fleet’, by L. He and G. Ma and W. Qi and X. Wang, 2018, Manufacturing and Service Operations Management, p. 28

mainly attributable to the declining charging infrastructure cost and the declining repositioning cost. A higher degree of centrality, e.g., $\xi = 1$ as Fig. 8b shows, would allow consolidating charging operations at the central district. Consequently, EV charging requires fewer charging sites and fewer chargers. The repositioning trips for charging are also shortened, as the charging locations overlap more with the customer trip origins and destinations.

The above result stresses the importance of charging operations management. Centrality causes supply/demand imbalances of the ride-sharing network. The platform has to use prices that are off the profit-maximizing level as an instrument to mitigate the imbalances. However, the relationship between centrality and the expenditure on rebalancing supply/demand is less significant.

4 Conclusion

It is nowadays imperative to bridge the gap between a smart-city future where shared and electrified mobility prevails versus a reality where it may actually fail. He et al. (2021) explicitly model customer’s endogenous usage of EVs at heterogeneous energy levels and the charging process using a charging up-to policy. In order to solve the nonlinear optimization problem, He et al. (2021) develop a computationally efficient lower- and upper-bound problems as MISOCPs. A series of numerical experiments are also conducted using a set of real operational data.

The case study on car2go’s operations in San Diego yields important findings. First, EV sharing infrastructure planners should concentrate limited charging resources at selected optimal locations. If it is a public charging network, it is crucial to ensure charger availability to the EV sharing fleet. Second, contrary to charging EVs only from energy levels around 20%, car2go should either proactively charge EVs from varying energy levels or adjust the threshold to 40%. Either policy change can enhance the EV utilization rate, improve the repositioning efficiency, and subsequently significantly increase the profit. Choosing the fleet size is also important. Third, if technology permits, higher charging power can alleviate the dependence on scarce charging resources, whereas larger battery capacity or unmanned repositioning can improve profitability. Finally, the study on urban spatial structure again highlights the importance of charging planning and operations to EV sharing.

There are still several research challenges that remain to address. For example, it will be desirable to develop adaptive decision making schemes for multi-stage infrastructure planning, together with charging and repositioning operations. In addition, the prevalence of shared and electrified mobility might cause problems to the future urban electrical grid. Also, a dynamic yet implementable charging policy (e.g., with time varying charge up-to levels) needs to be developed.

References

Abouee Mehrizi H, Baron O, Berman O, Chen D (2018) Adoption of electric vehicles in car sharing market. Working paper

Blomgren GE (2017) The development and future of lithium ion batteries. Journal of The Electrochemical Society 164(1):A5019–A5025

Boyacı B, Zografos KG, Geroliminis N (2015) An optimization framework for the development of efficient one-way car-sharing systems. European Journal of Operational Research 240(3):718–733

Desaulniers G, Errico F, Irnich S, Schneider M (2016) Exact algorithms for electric vehicle-routing problems with time windows. Operations Research 64(6):1388–1405

Fortune (2016) "Future cities could run on shared fleets of electric self-driving cars". http://fortune.com/2016/10/11/shared-electric-self-driving-cars/

Green L, Kolesar P (1991) The pointwise stationary approximation for queues with nonstationary arrivals. Management Science 37(1):84–97

He L, Mak HY, Rong Y, Shen ZJM (2017) Service region design for urban electric vehicle sharing systems. Manufacturing & Service Operations Management 19(2):309–327

He L, Ma G, Qi W, Wang X (2021) Charging electric vehicle sharing fleet. Manufacturing & Service Operations Management 23(2):471–487

Jiang D, Powell W (2016) Practicality of nested risk measures for dynamic electric vehicle charging. ArXiv e-prints 1605.02848

Kim SW, Mak HY, Olivares M, Rong Y (2018) Empirical investigation on the range anxiety for electric vehicles. Working paper

Lim MK, Mak HY, Rong Y (2015) Toward mass adoption of electric vehicles: Impact of the range and resale anxieties. Manufacturing & Service Operations Management 17(1):101–119

Mak HY, Rong Y, Shen ZJM (2013) Infrastructure planning for electric vehicles with battery swapping. Management Science 59(7):1557–1575

Schneider M, Stenger A, Goeke D (2014) The electric vehicle-routing problem with time windows and recharging stations. Transportation Science 48(4):500–520

Shortle JF, Thompson JM, Gross D, Harris CM (2018) Fundamentals of queueing theory, vol 399. John Wiley & Sons

Whitt W (1993) Approximations for the gi/g/m queue. Production and Operations Management 2(2):114–161

Yilmaz M, Krein PT (2013) Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles. IEEE Transactions on Power Electronics 28(5):2151–2169

Zhang H, Moura SJ, Hu Z, Qi W, Song Y (2017) A second order cone programming model for pev fast-charging station planning. IEEE Transactions on Power Systems Forthcoming

A Reactive Architectural Proposal for Fog/Edge Computing in the Internet of Things Paradigm with Application in Deep Learning

Óscar Belmonte-Fernández, Emilio Sansano-Sansano, Sergio Trilles, and Antonio Caballer-Miedes

1 Introduction

The amount of devices connecting to the internet has been growing exponentially in the last years, and projections talk about billions of devices connected by 2025 [1]. These devices will generate huge amounts of data streamed through the network at a high rate and with diverse formats [2]. Traditional data management systems will not be capable of handling these large data feeds instantaneously, with requirements known as the 3Vs: volume, variety and velocity [3].

Typically, these data produced by billions of heterogeneous devices have to be stored to perform analytics later or to feed other services. Some examples of these services can be location services based on Wi-Fi readings, geo-targeted advertising, weather forecasting services based on atmospheric data, urban planning or traffic management in the new era of the connected car, health care services such as clinical decision support systems or disease patterns analysis, etc.

These services are generally built upon artificial intelligence technologies such as machine learning (ML) and deep learning (DL) algorithms, applied to make reliable models for classification, regression or predictive tasks. These models should be adaptive and manage to modify their behaviour when the conditions change. In other words, to be useful, they must learn continuously from the incoming data and be able to generalize and accommodate their performance to new data. For example, a change in the traffic flow patterns can make the predictions of a model inaccurate.

Ó. Belmonte-Fernández (✉) · E. Sansano-Sansano · S. Trilles · A. Caballer-Miedes
Institute of New Imaging Technologies, Castellón de La Plana, Spain
e-mail: oscar.belmonte@uji.es; esansano@uji.es; strilles@uji.es; caballer@uji.es

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_9

Semi-supervised learning schemes can help to exploit vast amounts of unlabelled data to learn new data patterns mechanically and adapt trained models to data shifts and changes.

To train DL models in a reasonable time requires specialized hardware, such as modern graphics processor units (GPUs). This dedicated hardware is commonly provided as a service in the cloud, where DL models are trained and later used in tasks such as classification or regression problems. Although this approach presents lots of advantages, some issues arise when it is applied in the Internet of Things (IoT) realm, where not only cloud server but also any sensor in the wireless sensor network might fail, thus producing time delays in responses and loss of data. Fog/edge computing tries to alleviate these issues by approaching processing of data to the place where they are produced, addressing the following main issues: (1) on-time response, (2) resiliency in case of failure and (3) diverse workload management.

The objective of this work is to meet the fog/edge computing objectives building systems which are reactive, namely systems that are responsive, resilient and elastic, covering the latter three issues by the message-driven interchange between components of the system.

The remaining parts of this book chapter are organized as follows: Sect. 2 presents the characteristics of the IoT applications and how the new paradigm of fog/edge computing tackles the challenges in this field, DL is presented as the main IoT application relying on some intelligent modelling, and finally, the principles of reactive systems are presented. Section 3 presents our architectural proposal that maps reactive systems principles on top of the fog/edge computing paradigm to develop IoT systems. Two case studies are presented in Sect. 4 as a proof of concept of the proposed architecture. Conclusions are presented in Sect. 5.

2 Related Work

In this section, the IoT paradigm is first presented, showing some example applications where this paradigm has been successfully applied and highlighting its characteristics directly related to the reactive paradigm. Second, the fog/edge computing architecture is presented as an alternative to the cloud computing (CC); the main features directly connected with the reactive paradigm are highlighted as well. DL as an example that fits well with the requirements of reactive systems is presented in the third section. Finally, reactive systems are presented, and its characteristics, as stated in the reactive manifesto, are related to the IoT paradigm, fog/edge computing architecture and DL applications.

2.1 The Internet of Things

The IoT comprises things connected to the internet, with a unique identifier, that communicate, exchange data and, to some extent, are capable of performing some computation on the data they or others provide, alone or in collaboration with other connected things. Typically, IoT devices can sense the environment and interact with it using some attached actuators [4, 5, 6].

It is estimated that more than 100 billion devices will be connected to the internet by 2025 [1]. With this huge number of IoT devices continuously generating data in different formats, a significant concern is how to manage this enormous volume of data or, in other words, how to manage IoT big data [7].

Lightweight and straightforward messaging protocols such as message queuing telemetry transport (MQTT), extensible messaging and presence protocol (XMPP), constrained application protocol (CoAP) or advanced message queuing protocol (AMQP), among others [8], can play a central role in the IoT, allowing resource-constrained devices to interact effectively while providing fundamental services such as resource discovery, resource observation, security or quality of services (QoS) [4].

Figure 1 shows the four-layer architecture for common IoT applications. The sensing layer contains the physical devices in charge of measuring the variables of interest. The network layer is in charge of establishing communications between different things and with the infrastructures in the support layer. The service layer contains the computational logic to process the raw data coming from the sensor for generating valuable information. Applications using the information provided by the support layer are present in the application layer.

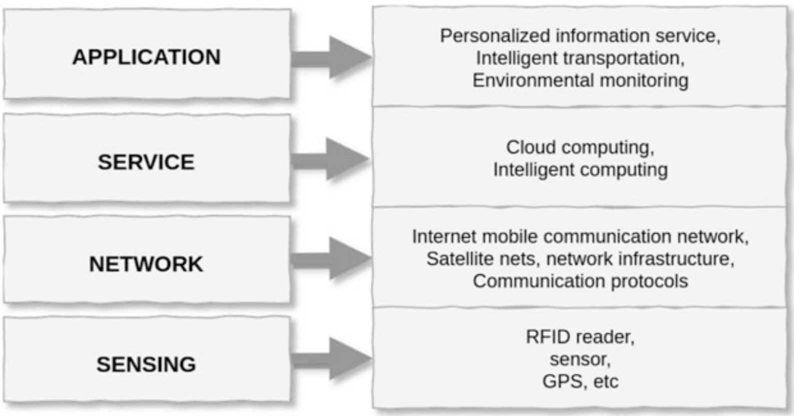


Fig. 1 The four-layer architecture for IoT

Security and privacy are major concerns in the IoT realm [9, 10, 11], especially in those applications managing personal data, as is the case with e-health and e-commerce, which handle sensitive data transmitted through the network and stored in remote databases. All four layers in IoT must be secured, and on the contrary, any security fault could compromise the whole system. One proposed solution to minimize security issues in IoT is to tackle the problem from a holistic perspective [12], where each layer is secured, taking into account its interactions with the other layers.

Interoperability is another critical concern in IoT due to the lack of standardization in the hardware, communications, protocols and logic layers [13]. One proposed solution for this lack of standardization is the use of open protocols and open hardware standards [14, 15, 16].

IoT has been successfully applied in some fields such as smart cities [17]. In [18] the SmartSantander project is presented, where IoT, Internet of Services (IoS) and internet of people (IoP) are combined as the building blocks for providing a smart city ecosystem. In [19], another implementation of the smart city paradigm in Padova, Italy, is presented, where the main objective is “to promote the early adoption of open data and information and communication technologies (ICT) solutions in the public administration”.

E-health is another field where the IoT paradigm has been taken as a base for developing new services. In [20], an e-health system addressed to diminish women mortality during pregnancy is presented. The population for the study was illiterate women in the rural part of developing countries. The solution proposed is based on the use of RFID tags, where health parameters such as blood pressure, blood sugar and others are stored and transmitted when a network node, located at rural health care centres, is near the RFID wore by the patient. A framework for bridging the gap between the expectations of elder people and technology products for health care is presented in [21]. This framework tries to minimize user refusal and maximize the quality of monitoring from a health care perspective. The work presented in [22] is addressed to monitor elder people with mild cognitive impairment and frailty. For tracking the elder behaviour, data coming from the user, his/her home, and the environment, is used, which in turns is the basis for identifying risks.

Wireless sensor networks are used in [23] for long-term environmental monitoring and as a basis for IoT applications. The deployed sensor nodes are low cost, energy efficient and with low maintenance. The solution proposed is validated by deploying 1000 sensors for monitoring birdhouses on the field. The authors in [15] developed a general-purpose environmental monitoring system based on open hardware and open standards to minimize interoperability issues. The solution proposed is flexible enough to include new sensors with low effort. Energy consumption issues are minimized by providing each thing on the internet with a LiPo battery attached to a solar panel to guarantee long running of the solution.

2.2 Fog/Edge Computing

CC provides computing services and data storage over the internet, with the benefits of flexibility and efficiency. Emerging systems and applications, such as IoT, cyber-physical systems (CPS) [24, 25, 26], fifth-generation networks (5G) [27] and embedded artificial intelligence (EAI), will produce massive amounts of data, imposing a big challenge on the current CC architecture.

The main challenges that should be addressed to meet the requirements these emerging technologies impose are as follows:

- **Increased network traffic load**. The growing number of connected devices will generate vast amounts of data. Sending these massive volumes of data to remote servers in the cloud would require high network bandwidth, not to speak about energy consumption and network management costs.
- **Increased latency**. IoT applications and systems such as smart grid, smart cities or critical safety services, require real-time responses with end-to-end latencies between the sensor and the control node within a few milliseconds to function properly [28]. With the current cloud architecture, a request to/from cloud may take several milliseconds to seconds to travel from client to cloud service provider [29]. The high load that IoT systems would impose on cloud's central servers would increase network overhead, resulting in slow response time, since such requirements fall outside what current cloud services can achieve.
- **Increased resources**. Battery life and computing capabilities are scarce resources for IoT devices such as sensors, actuators, controllers, surveillance cameras, embedded medical devices, etc. Such devices are not able to rely on their own constrained resources to satisfy their computing needs and cannot interact directly with cloud servers since this would demand resource-intensive processing and complex protocols.
- **Scalability and fault tolerance**. The goal of fault tolerance in IoT is to better adapt to changing environments and build up trustworthy redundancy since users expect their applications and devices to operate correctly and continuously. However, in real IoT deployment scenarios, systems are composed of multiple devices that have different hardware characteristics, have different failure properties and are deployed with different purposes and capabilities. This heterogeneity and constant evolution of IoT systems pose a big challenge to building up redundancies and adapting to changing environments.

To meet these challenges, fog/edge computing [30, 31, 32, 33, 34, 35, 36] proposes an extension to CC to transition from big and distant cloud servers to smaller computational devices closer to users and sensors. Fog/edge computing provides computing, storage, control and networking services for customers or applications at the edge of the network and the space between networking central servers and end-users. It is designed as a distributed architecture that allows the data generated by IoT devices to be processed and stored in networking edge devices and smart clients, suitable to provide efficient and secure services for a large number of end-users. Fog/edge architectural principles are designed to overcome the inadequacy of the traditional cloud and offer a new scenario for the IoT requirements. The predictions are that by the year 2019, 45% of the data generated by IoT will make use of edge computing [37].

The fog/edge computing paradigm includes techniques such as fog computing [30], cloudlets [38], mobile edge computing and microdata centres to implement a middle layer between the end-user and cloud resources, with the aim to minimize both response delay and network traffic, which are two of the significant constraints in the cloud paradigm. Regarding network traffic, it has been shown that caching at the edge of the network considerably reduces access latency and network traffic [39]. Fog/edge technologies make use of dedicated services at the edge to provide real-time response and data filtration. Downward network traffic can be decreased by servicing users at network edge instead of using remote cloud data centres, while upward traffic can be lessened by preprocessing and filtering raw data and uploading the processed data. Concerning the latency, content deployment at the network edge is critical to decreasing the response time. Bringing CC to the proximity of end-users and, thus, minimizing network traffic and latency lead to energy efficiency and data costs reductions.

Since there are not standardized definitions yet, many edge technologies are being proposed to implement fog/edge computing, such as mobile edge computing (MEC), microdata centres (MDCs) or cloudlets. All these technologies are designed to work in cooperation to give support to enable applications such as pervasive health care, smart cities, smart grid, etc. To fulfil the real-time requirement of these applications, these technologies will complement and extend cloud services to the edges of the network, increasing the responsiveness of the system.

To address scalability and fault tolerance issues, the fog/edge architecture takes advantage of the existing redundancy in case of failure, having the devices to collaborate with one another and to be sharable among different applications, using multiple types of networks and any connections that are available. Such IoT devices should be able to both communicate via access points of the infrastructure if they are available and interact with one another autonomously, without the infrastructure as an intermediary if the latter is not available. In such scenario, each IoT device behaves as both host and router.

Edge computing enables the offloading of energy-consuming application from resource-constrained IoT devices to the fog/edge nodes, trying to find an optimal trade-off between the computational energy consumption and the transmission energy consumption. If the edge has computational resources, it provides a chance to offload part of the workload both from end devices, liberating them from computation tasks, and from cloud servers, diminishing the energy cost of the transmission.

As regards the computational load balance, the end devices can offload some of their high processing tasks to the nearby edge devices to reduce their load, or, if the end device is capable, perform some form of data filtering to discard the redundant data, so only filtered data is sent to the cloud. On the energy cost side, studies on the energy cost of internet data transportation show that 14% of the energy consumption is due to data transportation [40]. Part of this energy can be saved by employing intelligent data caching techniques at edge devices to reduce the burden on the core network and optimizing the synchronization frequency of data between edge devices and cloud servers [41].

To enhance the robustness of an IoT architecture, service-oriented architecture (SOA) provides IoT with an abstraction of combinable and manageable services that facilitates the cooperation of different devices to achieve fault tolerance. When a fault happens to a device, this architecture can reconfigure the system by using devices of other modalities to cover the fault, addressing the problem in three different stages of service management: service provider, service registry and service client [42].

2.3 Deep Learning

DL has achieved many advances in recent years, dramatically improving the state of the art in many artificial intelligence tasks like computer vision, language processing, speech recognition and many more. With its deep architecture, composed of multiple non-linear hidden layers, deep learning models can learn very complicated relationships between their inputs and outputs and thus solve complicated AI tasks.

Although the theoretical basis for deep learning algorithms has been around since the second part of the twentieth century, there are three main reasons for the present boom of this ML field:

- **More algorithms**. Since 1943 McCulloch and Pitts computational neuron model [43], which was the first attempt of simulating the operation of a neuron as a computational tool, researchers have made many significant advances proposing new neural network architectures, learning algorithms and training procedures. Several deep learning achievements have been possible thanks to new neural network models such as convolutional neural networks (CNN) [44, 45], deep belief networks (DBN) [46], deep auto-encoders (AE) [47] or recurrent neural networks (RNN) [48], among many others [49]. Along with the proposal of new models, researchers have introduced new ways of training neural networks more efficiently, in terms of speed, accuracy and generalization capability, with new techniques such as dropout [50] or L1 and L2 regularization [51], which helps to prevent over-fitting, new gradient methods as RProp [52] or Adagrad [53], and recent optimization algorithms such as batch normalization [54] or layer normalization [55].
- **More computational power**. The limitations of shallow neural network architectures [56] make it necessary to build deeper architectures to increment the expressive power of networks, and more representation power requires more computation complexity. Today DL models are composed of billions of parameters, and training these models requires both high bandwidth and computation power. This scale of complexity and amount of data demands

scalable computation resources, and distributed architectures can achieve this. Scalable communication architectures are being proposed [57, 58] to efficiently distribute the training process among distributed GPU clusters. These architectures aim to address the inefficiencies caused by parameter synchronization between clusters via the network since communications usually bottleneck the training computations.

- **More data**. DL algorithms work significantly better and are more useful when the amount of available training data increases. A rough rule of thumb is that a supervised DL algorithm can achieve acceptable performance with around 5000 labelled examples per category and can match human performance when trained with a dataset containing at least 10 million labelled examples. Even though the first artificial neural networks were proposed in the 1950s, DL has only recently become recognized as a crucial technology, since they can now be trained with the resources they need to succeed. The size of available datasets has expanded remarkably over time, thanks to the increasing digitization of society. As IoT technologies develop, huge amounts of new data will be available to train more complex models that can reach and exceed human performance on complex tasks.

  When applications do not require instant real-time data updates, storing sensor data and sending aggregate data bursts can reduce the number of transmissions. Moreover, replication and distributed storage techniques [59, 60] can be implemented to reduce the probability of data loss and network bandwidth needs, getting nodes to cooperate by storing the sensed data in a distributed way.

One of the core advantages of DL is its ability to learn features from raw data automatically. In the ML field, finding a good set of features from the raw data to isolate key information and highlight important patterns is crucial to make algorithms work, but this task requires expert knowledge and is difficult and time-consuming. The use of feature engineering to improve the effectiveness of ML predictive models is both laborious and expensive and requires high-level expertise. DL algorithms obviate the need for hand-crafted feature extraction by automating the process of discovering the best features directly from the available data.

The use of DL algorithms is thus especially suited for IoT since the amount of available data is vast. But, generally, the data produced by sensors is unlabelled, and obtaining label information is difficult, expensive or impractical. Using semi-supervised learning techniques to make maximum use of this data allows for effective generalization from small labelled datasets to large unlabelled ones. DL semi-supervised algorithms such as deep generative models [61], one-shot learning [62], ladder networks [63] or semi-supervised embedding [64] can be implemented to improve model accuracy in architectures that cope with large amounts of unlabelled data.

2.4 Reactive Systems

The fast-growing of users' expectations on how applications behave has boosted the creation of new architectural paradigms to meet such expectations. Many applications, such as virtual and augmented reality, pervasive health care [65], ubiquitous cognitive assistance [66, 67], IoT, interconnected vehicles and smart cities, to cite a few, need to minimize time delays of the service regardless of the number of clients connected and in any adversary circumstances [68].

As presented in Sect. 2.2, these are the issues that fog/edge computing faces. In this chapter, we propose to develop such systems from the new paradigm of reactive systems, which tries to fit users' expectations, such as pervasive applications running on mobile, desktop and in a browser, with the afore presented expectations.

As stated in the *reactive manifesto* [69], reactive systems are

- responsive: the system responds promptly if at all possible;
- resilient: the system stays responsive in the face of failure;
- elastic: the system stays responsive under a varying workload;
- message driven: asynchronous message passing between system components.

The systems developed following these principles can attend a significant number of requests per second, can rescale depending on the workload and minimize the issues in case of failure, while orchestrating all requirements using asynchronous decentralized message passing between different actors in the system. Also, message-driven communication allows location transparency; different actors in the system do not know the address where their requests will be finally attended, just the name of the resource that will attend it, allowing resiliency by isolating and self-healing of the resources.

Reactive programming [70] plays a central role when developing reactive systems. Reactive message-driven applications are based on the asynchronous interchange of messages between different components in the system, where the observer/observable design pattern is extended in such a way that components in the application can express their interest in being notified any time a value change in another component. Most popular programming languages, such as Java, JavaScript and .Net, to cite a few, have extensions for developing reactive applications.

3 Architecture Proposal

This section presents first the proposed architecture based on the principles of reactive systems to fulfil the requirements of the fog/edge computing. Then, a proof of concept in the realm of e-health is presented showing how the generic proposed architecture can be implemented in a particular case.

3.1 Reactive Systems and Fog/Edge Computing

The availability of storage, computing and platform services that reside on remote servers and are accessible through the internet, the so-called CC paradigm, has revolutionized the way users use these services in recent years. However, some IoT requirements are not entirely fulfilled by this paradigm, in particular, bandwidth efficiency, context awareness and mobility support, among others.

Fog/edge computing aims at addressing these issues by approaching storage and computation services to the edge, namely, as close to the point where they are produced as possible. This new paradigm is not a complete replacement of the CC paradigm; on the contrary, some services can be approached to the thing where they are produced, like storage or basic computations, but other can be only achieved in the CC realm.

One example where a collaboration of both paradigms can be seen as a win–win approach is the use of DL models for the prediction on massive data. To train DL models is a time-consuming task that requires the use of specialized hardware to be performed in a reasonable time. Once such a model has been trained, the computational power to use them for making predictions does not require as many resources. Namely, data coming from the fog/edge can be used to train DL models in the cloud, and once trained, the models can be downloaded to the edge/cloud, where each resource must be resilient, isolated and self-healing to quickly recover from any fault, thus promoting the resiliency of the whole system and its elastic scalability.

From our point of view, the characteristics mentioned above can be entirely covered by the previously presented principles of reactive systems (see Sect. 2.4). This paradigm proposes systems that respond promptly (responsiveness), are robust in case of failure (resiliency), manages varying workload efficiently (elasticity), and the components in the system interact with each other by asynchronous message-passing (message-driven). Note that there is no distinction in reactive systems between fog/edge or cloud components, both have the same requirement in order to make the whole system reactive. Reactive programming can be very valuable when developing systems with such requirements; any component in the system can get the role of an observable or an observer, or both, reacting only when messages indicating changes in other components of its interest happen, in an asynchronous decoupled way.

3.2 Our Proposal of Reactive Deep Learning Systems for the IoT

Figure 2 shows the four-layer architecture proposed in this work for building reactive fog/edge systems for IoT. The main building block of this architecture is the sensor, meaning by that a physical device using a wireless or a cable connection

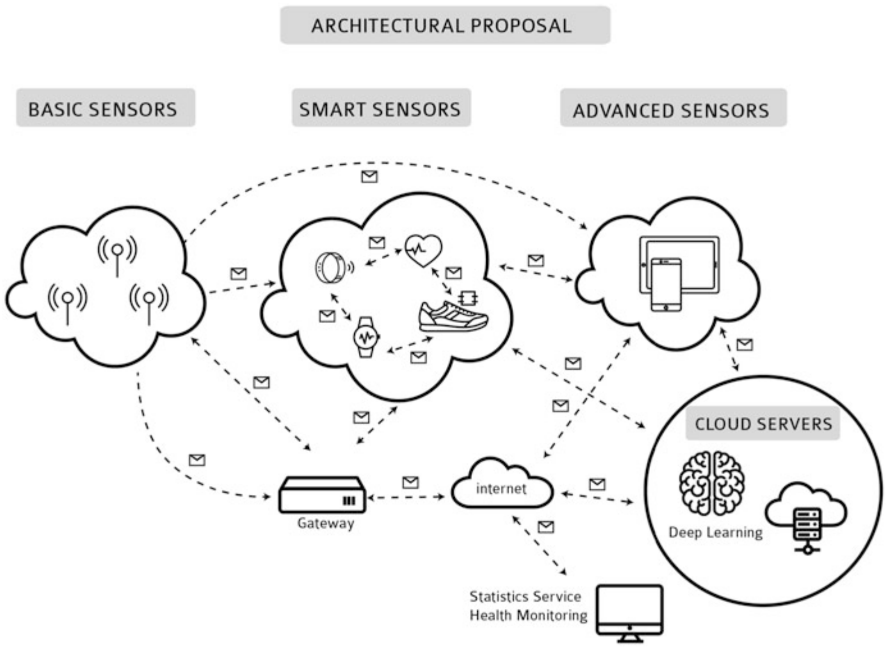


Fig. 2 Proposed architecture

to interchange messages with other sensors, gateways or servers and that can sense or/and interact with its surroundings.

The sensor is autonomous, energy efficient and ubiquitous and cooperates with other sensors to provide reliable information related, not only to what it can sense or interact with but also to other sensors. In case of failure, they can be considered self-organized to some extent.

A sensor, depending on its computational power and storage capabilities, can delegate some services to another sensor or to a gateway and even can completely share the services it provides with other federated sensors in order to minimize response degradation in case of failures. We define three types of sensors based on their computational capabilities: base, smart and advanced sensors. Figure 3 shows the interaction between sensors with the four-layer IoT architecture. Basic sensors, such as temperature sensors, can sense its surroundings and send the measurements to a gateway node. Smart sensors are also able to store and perform some calculations on the measurements. They can be requested by clients, which can be another smart sensor, to obtain the plain measurements or the result of basic calculations. Finally, advanced sensors can also execute applications running in system-on-a-chip (SoC) hardware, which can be remotely deployed on it. Smart and advanced sensor can replicate data of its nearest neighbours to increase the responsiveness of the system and to maintain the performance of the whole system in case of failures.

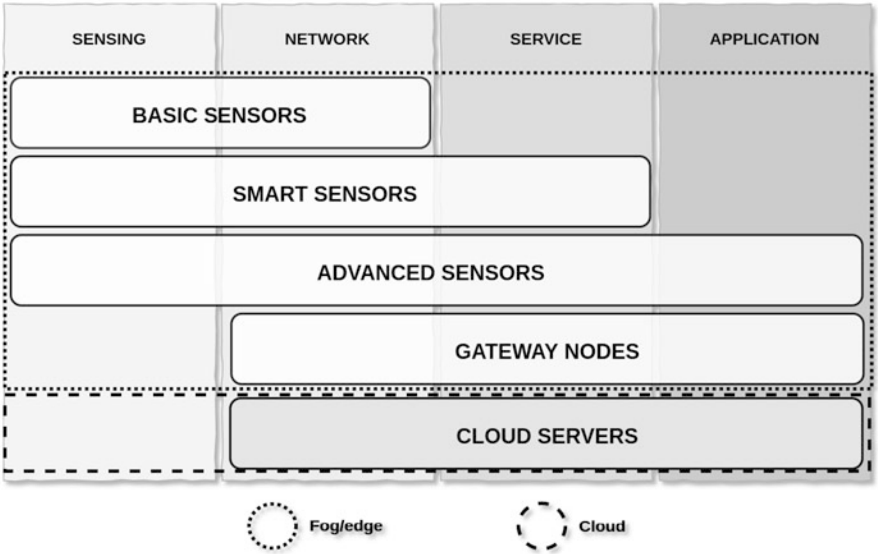


Fig. 3 The four-layer architecture for IoT

Gateway nodes are in charge of storing and replicating data from sensor of any type. Moreover, they are bridges between basic sensors and cloud servers. They can act as sensor proxies to timely serve request in case of an increasing workload, guaranteeing the elastic behaviour of the system.

Sensors and gateways are a realization of the fog/edge computing paradigm providing storage, computation and networking services between them and cloud servers [31]. Cloud servers are not removed from the fog/edge paradigm; instead, they coexist with the latter to reach reactive functionality.

On the network layer, TCP/IP is the base protocol for message interchange between different things. Above this protocol, other message-oriented, lightweight and publish/subscribe protocols can be used, such as the already mentioned MQTT protocol, being this the nervous system of the proposed architecture. Such protocols are used to interchange messages between the different actors in the architecture and to guarantee the quality of service of delivered messages.

Fog/edge computing and CC coexist in the IoT paradigm. Instead of being competitive paradigms, they collaborate and complement each other. CC is well suited for supporting high computational demanding services as training deep neural networks, which currently rely on specialized hardware, consuming lots of energy and computational resources. These high demanding requirements are challenging to provide by sensors and gateways, while CC can cope with them with no difficulty. Moreover, the principles of reactive systems can be easily achieved by the CC paradigm, and in nature they respond to client requests promptly, can cope with failures by replicating the services they provide and also can deal with varying workload by creating or removing service instances.

DL models are expensive to build regarding hardware resources, computing time and energy consumption. Specialized hardware, such as GPU, dramatically reduces the computational time at the cost of big amounts of energy consumption. CC infrastructures are well suited for efficiently managing high demanding resources, using techniques such as charge balancers and geographical concentration. DL models can be built in the cloud servers with the data provided by sensor and gateways in the fog/edge. Once trained, these models can be delivered back to gateways and sensors with enough power to run the built DL model. In the same way, the incoming data can be used with semi-supervised techniques to improve previously trained models.

4 Case Studies: Active and Healthy Ageing and Smart Farming

In this section, we present a couple of use case examples to illustrate two particular implementations of the proposed architecture. These examples describe an IoT architecture for active and healthy ageing (AHA) and smart farming.

4.1 Active and Healthy Ageing Scenario

AHA is a major societal challenge, common to all populations that we will be facing in the upcoming decades. The word active refers not just to the ability to be physically active, but to continuing participation in social, economic, cultural, spiritual and civic affairs. Health refers to the physical, mental and social well-being. AHA allows people to realize their potential for physical, social (economic, cultural, spiritual and civic affairs) and psychological well-being throughout the life course.

Demographic ageing profoundly impacts on society and economy. The ratio of working people versus inactive people will shift from 4:1 to 2:1 in 2050, and a shortage of up to 2 million jobs in care and health is projected to emerge already by 2025 [71]. Innovating technologies can help turn the ageing challenge into an opportunity, by improving the quality of life for older citizens, allowing more efficient and sustainable care systems, and enabling new market opportunities and economic growth.

The goal of the proposed IoT infrastructure is to provide services such as human activity recognition [72], context modelling (user-profiles and preferences), anomaly detection, location and identity identification (PIR sensors) and planning, to promote active ageing on older adults. The proposed system will be composed of multiple types of sensor devices, such as in-home devices (presence detectors, microphones, pressure detectors, video cameras, temperature sensors, humidity sensors, air quality sensors), smart wearables (smartwatches, smart glasses, fitness bands), tablets, smartphones, etc. Depending on their computational capabilities, the network status, the system configuration or the operation conditions, some of these devices can work either collaboratively or autonomously, can either connect to a local node to send data and perform some analysis or act as a distributed system by coordinating their actions by passing messages.

Unlike current cloud-only architectural approaches, the proposed architecture can sustain the projected data velocity and volume requirements of the IoT by auto-adjusting its behaviour dynamically from the environmental conditions. Real-time monitoring and detection of anomalies (activity irregularities, the fall of an elderly citizen, heart attacks) pose low latency requirements on AHA systems from the standpoint of both detection and response.

To illustrate this, consider the next example: a 60-year-old woman with a heart condition lives alone on the fifth floor of a building. Her home is equipped with simple sensors such as PIRs, which detect her presence in a room. She has a pacemaker or an implantable cardioverter defibrillator (ICD) implanted and wears an activity tracker on her wrist. Additionally, she has a smartphone and a tablet that, besides their main function, are part of the framework and continuously collect and analyse data. Ordinary and non-intrusive devices collect all the data. The installed sensors do not interfere in the daily life of the user, and the system is designed to be non-invasive and not to expect any help or effort from the end-user, keeping the user interaction with the system intuitive and simple.

The system (see Fig. 4) is deployed in conjunction with a fog/edge node (or IoT gateway) that processes the incoming data from the sensors. This architecture

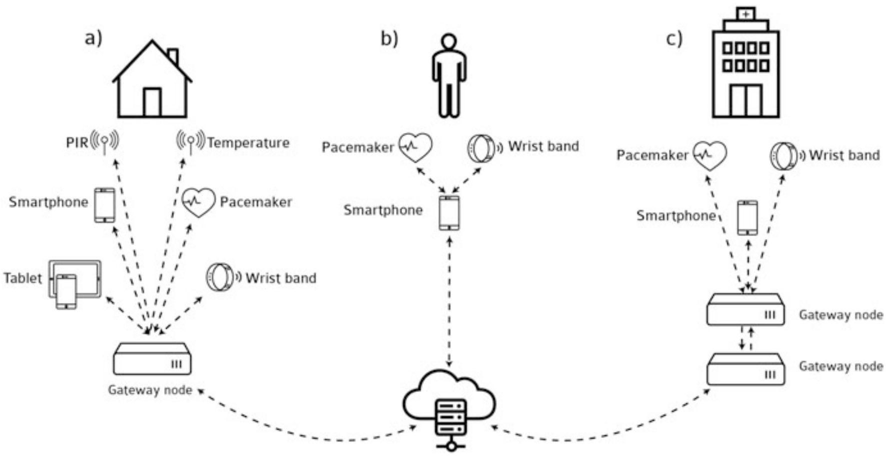


Fig. 4 Active and healthy ageing study for the proposed architecture. (**a**) Sensors and gateway located at user's home. (**b**) Sensors worn by the user. (**c**) Sensors and gateways located in a hospital

allows placing intelligence and processing capabilities closer to where the data originates, with the aim to reduce the amount of data sent to the cloud and thus decrease network and internet latency, improving system response time. This node is designed to connect to the cloud servers to synchronize data, add, update or remove available services, update predictive models, perform compute-intensive tasks, etc. When the user leaves its home, her smartphone or her tablet can become a mobile fog/edge node that communicates with the cloud or with other nodes as they become available. This mobile node can also be capable of performing some operations autonomously in circumstances where it cannot connect to other fog/edge nodes or the cloud, receiving readings from the available sensors (activity tracker, pacemaker, smartphone inertial sensors) and providing basic services to the user. For example, her wearable inertial sensors continue to sending their readings, and the mobile node detects if she is using the elevator, the stairs, public transportation, etc. The stored operational history can be aggregated and sent later to the cloud to produce large-scale analytics and create optimized models.

This particular architecture can be easily extended to nursing homes, hospitals, assisted residences and apartments for elders or disabled people, etc., with some gateway nodes depending on the number of users and sensing devices, and the workload and responsibilities of each node adapting dynamically to the particular operating conditions. Reliability and resiliency are compromised if the cloud or a single place in the system's hierarchy is the only location where decisions can be made. The described deployment continues to deliver functionality under adverse operating conditions. If cloud servers or local nodes are not available, intelligent devices can assume part of their functions and continue providing services to the user, safeguarding the availability and integrity of data, and compute on edge nodes using enhanced hardware, software and network designs.

The application layer of this particular example is distributed between the central cloud servers, the fog/edge nodes and the smart devices. The main servers can use DL algorithms to generate predictions from the data acquired by the sensors and to enhance its learned models through a semi-supervised learning implementation. The nodes and smart devices will use these models to provide services to other devices and users. An example of such service could be a reward system based on user activity. If the user achieves some goals that have beneficial effects on her health, the system can use available devices, such as the tablet or the smartphone, to give positive reinforcement and encourage the user to keep improving. On the contrary, the system can provide negative feedback and show warnings if the user actions may pose a potential health risk.

4.2 Smart Farming Scenario

The second scenario proposed focuses on the context of precision agriculture or smart farming [73]. The advance in technology for environmental monitoring has also extended into agriculture and farming [74]. Highly accurate embedded environmental sensors have paved the way for *precision agriculture*, which applies more efficient irrigation, targeted use of fertilizers and pesticides and more precise use of fodder and antibiotics for livestock. This enhanced form of farming potentially leads to increased productivity, greater yields and a reduced environmental footprint. Taking precision farming even further in an IoT context, the concept of *smart farming* has emerged for decision support for farmers. Smart farming focuses on real-time data gathering, processing and analysis and the automation of farming procedures for an overall improvement of farming operations and management [74].

More specifically, this scenario is focused on the cultivation of vineyard, to increase the production quality and yield [75]. Viticulture is a field in which a significant adaptation of IoT platforms has been performed. It has historically been characterized by the high-quality product that can be obtained. The winegrowing can be affected by several factors: the varietal wines, the selection of climatic zones and suitable soil and the winegrower practices in managing the vineyards [76].

The main goal of this scenario is to cover the whole workflow, from the production in fields until its distribution using agricultural cooperatives. Three different contexts or actors are defined: *field*, *farmer* and *cooperative* (Fig. 5). Each context has and requires different features and configurations, depending on each functionality. At the configuration level, a sensorial skin of IoT devices [16] is deployed, different nodes with meteorological sensors are available on the field, such as temperature, soil and air humidity, precipitation and wind speed and direction. Other sensors and devices could be added, such as air quality sensors or even drones with thermal or multispectral cameras or add-ons to realize fertilization or planting tasks. In addition, other devices such as tractors with IoT connectivity

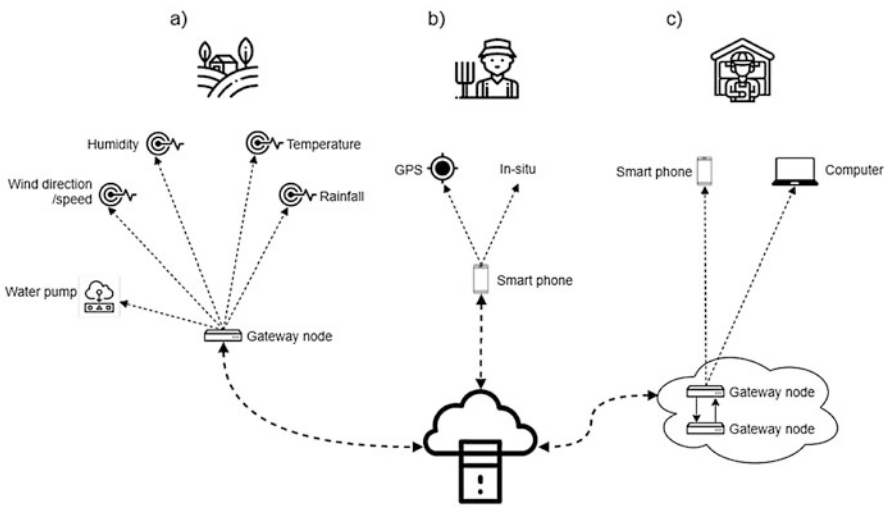


Fig. 5 Smart farming study for the proposed architecture. (**a**) Sensors and gateway located in the field. (**b**) Sensors worn by the farmer. (**c**) Sensors located in the cooperative building

could be present for tilling or fertilization tasks. Moreover, also capable actuators are added to perform actions within the same plots of vines, for example, pumps or automatic irrigation. These actuators will be activated from the fog nodes.

A vineyard farmer has different vineyard smallholdings. In each vineyard smallholding, a weather station with internet connectivity will be deployed. Each weather station is composed of different nodes, such as: temperature sensor nodes, soil and air humidity sensor nodes, wind direction and speed sensor nodes and rain gauge sensor nodes. These nodes with fog/edge (or IoT gateway) features can analyse incoming data from each environmental basic sensor. These nodes allow to follow models to detect vineyard diseases [77, 78, 79, 80], for instance, *black rot*, *botrytis*, *powdery mildew* or *downy mildew*. All these diseases are contingent on certain environmental conditions. In the same way as in the previous scenario, this architecture reduces the amount of data sent to the cloud and network and internet latency. Based on the results of the analyses, different actions will be triggered, such as starting irrigation to apply some type of treatment or launching alerts at the agricultural and cooperative level. The first one will be in charge of performing the validations and treatments manually and make the report of the practices carried out and observations obtained through their mobile phone. On the other hand, the cooperative will be in charge of advising farmers of similar or nearby plots where the disease has been detected and will also apply the necessary measures to mitigate the damage to the product.

Based on the results of the analyses, different actions will be triggered, such as starting irrigation to apply some treatment or launching alerts to the farmer or the cooperative. The first one will be in charge of carrying out the validations and treatments manually and make the report of the practices carried out and observations obtained through his/her smartphone. On the other hand, the cooperative will be in charge to advise other farmers with similar or nearby parcels where the disease has been detected and will also apply the necessary measures to mitigate product damage.

Based on the results obtained from previous analysis and data collected by sensors, the central servers can use deep learning techniques to generate predictions and prevent farmers and cooperative managers. Some indicators can be offered, such as quantity and types of plant protection products that they should buy during a campaign, best practices to prevent or mitigate diseases or the price that they can sell the product to an end-user in the form of a wine bottle.

5 Conclusions

A reactive architectural proposal for building fog/edge computing systems in the IoT paradigm and its application to DL computation is presented in this work.

This proposal aims at providing a robust solution for current high demanding applications, particularly in the IoT realm. Low latency, high throughput, information updated and device diversity are demanding characteristics of nowadays applications. On the one hand, the new fog/edge computing paradigm approaches computation, storage, network management and services to capable devices; on the other hand, fog/edge computing is not a substitution of the CC paradigm. Instead, both paradigms can coexist and seamlessly collaborate guided by the principles of reactive systems.

Reactive systems are addressed to cope with the high-demanding characteristics of current applications based on four principles: responsiveness, resiliency, elasticity and message-driven. From our point of view, these principles can guide the development of fog/edge computing, especially in the case of the IoT applications.

Two scenarios are proposed as a proof of concept. We show how the proposed architecture can be particularized in the case of an e-health application, where different computing levels are required depending on the geographical location of the user. The second scenario applies the architecture in the context of smart farming in order to prevent and to mitigate the effect of diseases in vineyard smallholdings.

Acknowledgments This work has been partially funded by the Spanish Ministry of Science, Innovation and Universities through the "Retos investigación" programme (RTI2018-095168-B-C53), by Jaume I University "Pla de promoció de la investigació 2020" programme (UJI-B2020-36) and by Generalitat Valenciana "Promoció de la investigació 2020" programme (AICO/2020/046). Óscar Belmonte-Fernández had a grant by the Spanish Ministry of Science, Innovation and Universities (PRX18/00123) for developing part of this work. Sergio Trilles has been funded by the Juan de la Cierva - Incorporación postdoctoral programme of the Ministry of Science and Innovation - Spanish government (IJC2018-035017-I).

Conflicts of Interest The authors declare that there are no conflicts of interest regarding the publication of this chapter.

References

1. R. Taylor, D. Baron, D. Schmidt, in *2015 10th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT)* (IEEE, 2015), pp. 192–195
2. S. Trilles, Òscar Belmonte, S. Schade, J. Huerta, International Journal of Digital Earth **10**(1), 103 (2017). https://doi.org/10.1080/17538947.2016.1209583
3. S. Sagiroglu, D. Sinanc, in *2013 International Conference on Collaboration Technologies and Systems (CTS)* (2013), pp. 42–47. https://doi.org/10.1109/CTS.2013.6567202
4. L. Atzori, A. Iera, G. Morabito, Computer Networks **54**(15), 2787 (2010). https://doi.org/ 10.1016/j.comnet.2010.05.010. http://www.sciencedirect.com/science/article/pii/S1389128610001568
5. J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Future Generation Computer Systems **29**(7), 1645 (2013). https://doi.org/10.1016/j.future.2013.01.010. http://www.sciencedirect.com/science/article/pii/S0167739X13000241
6. A. Bahga, V. Madisetti, *Internet of Things: A hands-on approach* (VPT, 2014)
7. A. Gandomi, M. Haider, International Journal of Information Management **35**(2), 137 (2015). https://doi.org/10.1016/j.ijinfomgt.2014.10.007. http://www.sciencedirect.com/science/article/pii/S0268401214001066

8. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, IEEE Communications Surveys Tutorials **17**(4), 2347 (2015). https://doi.org/10.1109/COMST.2015.2444095
9. R. Roman, P. Najera, J. Lopez, Computer **44**(9), 51 (2011). https://doi.org/10.1109/MC.2011.291
10. H. Suo, J. Wan, C. Zou, J. Liu, in *2012 International Conference on Computer Science and Electronics Engineering*, vol. 3 (2012), vol. 3, pp. 648–651. https://doi.org/10.1109/ICCSEE.2012.373
11. Q. Jing, A.V. Vasilakos, J. Wan, J. Lu, D. Qiu, Wireless Networks **20**(8), 2481 (2014). https://doi.org/10.1007/s11276-014-0761-7
12. R. Roman, J. Lopez, M. Mambo, Future Generation Computer Systems **78**, 680 (2018). https://doi.org/10.1016/j.future.2016.11.009. http://www.sciencedirect.com/science/article/pii/S0167739X16305635
13. D. Bandyopadhyay, J. Sen, Wireless Personal Communications **58**(1), 49 (2011). https://doi.org/10.1007/s11277-011-0288-5
14. S. Trilles, O. Belmonte, L. Díaz, J. Huerta, IEEE Sensors Journal **14**(12), 4143 (2014). https://doi.org/10.1109/JSEN.2014.2339931
15. S. Trilles, A. Luján, Ó. Belmonte, R. Montoliu, J. Torres-Sospedra, J. Huerta, Sensors **15**(3), 5555 (2015)
16. C. Granell, A. Kamilaris, A. Kotsev, F.O. Ostermann, S. Trilles, in *Manual of Digital Earth* (Springer, 2020), pp. 387–423
17. S. Trilles, A. Calia, Ó. Belmonte, J. Torres-Sospedra, R. Montoliu, J. Huerta, Future Generation Computer Systems **76**, 221 (2017)
18. J.M. Hernández-Muñoz, J.B. Vercher, L. Muñoz, J.A. Galache, M. Presser, L.A.H. Gómez, J. Pettersson, in *The Future Internet Assembly* (Springer, 2011), pp. 447–462
19. A. Zanella, N. Bui, A. Castellani, L. Vangelista, M. Zorzi, IEEE Internet of Things Journal **1**(1), 22 (2014). https://doi.org/10.1109/JIOT.2014.2306328
20. V.M. Rohokale, N.R. Prasad, R. Prasad, in *2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE)* (2011), pp. 1–6. https://doi.org/10.1109/ WIRELESSVITAE.2011.5940920
21. A. Soro, A.H. Ambe, M. Brereton, Wireless Communications and Mobile Computing **2017** (2017)
22. A. Almeida, A. Fiore, L. Mainetti, R. Mulero, L. Patrono, P. Rametta, Wireless Communications and Mobile Computing **2017** (2017)
23. M.T. Lazarescu, IEEE Journal on Emerging and Selected Topics in Circuits and Systems **3**(1), 45 (2013). https://doi.org/10.1109/JETCAS.2013.2243032
24. A.A. Cardenas, S. Amin, S. Sastry, in *2008 The 28th International Conference on Distributed Computing Systems Workshops* (2008), pp. 495–500. https://doi.org/10.1109/ICDCS. Workshops.2008.40
25. F.J. Wu, Y.F. Kao, Y.C. Tseng, Pervasive Mob. Comput. **7**(4), 397 (2011). https://doi.org/10.1016/j.pmcj.2011.03.003
26. S.H. Ahmed, G. Kim, D. Kim, in *2013 IFIP Wireless Days (WD)* (2013), pp. 1–5. https://doi.org/10.1109/WD.2013.6686528
27. R.N. Mitra, D.P. Agrawal, ICT Express **1**(3), 132 (2015). https://doi.org/10.1016/j.icte.2016.01.003. http://www.sciencedirect.com/science/article/pii/S2405959515300503. Special Issue on Next Generation (5G/6G) Mobile Communications
28. M. Weiner, M. Jorgovanovic, A. Sahai, B. Nikolié, in *2014 IEEE International Conference on Communications (ICC)* (2014), pp. 3829–3835. https://doi.org/10.1109/ICC.2014.6883918
29. K. Ha, P. Pillai, G. Lewis, S. Simanta, S. Clinch, N. Davies, M. Satyanarayanan, in *2013 IEEE International Conference on Cloud Engineering (IC2E)* (2013), pp. 166–176. https://doi.org/10.1109/IC2E.2013.17
30. F. Bonomi, R. Milito, J. Zhu, S. Addepalli, in *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing* (ACM, New York, NY, USA, 2012), MCC '12, pp. 13–16. https://doi.acm.org/10.1145/2342509.2342513

31. F. Bonomi, R. Milito, P. Natarajan, J. Zhu, *Fog Computing: A Platform for Internet of Things and Analytics* (Springer International Publishing, Cham, 2014), pp. 169–186
32. L.M. Vaquero, L. Rodero-Merino, SIGCOMM Comput. Commun. Rev. **44**(5), 27 (2014). https://doi.acm.org/10.1145/2677046.2677052
33. S. Yi, C. Li, Q. Li, in *Proceedings of the 2015 Workshop on Mobile Big Data* (ACM, New York, NY, USA, 2015), Mobidata '15, pp. 37–42. https://doi.acm.org/10.1145/2757384.2757397
34. W. Shi, J. Cao, Q. Zhang, Y. Li, L. Xu, IEEE Internet of Things Journal **3**(5), 637 (2016). https://doi.org/10.1109/JIOT.2016.2579198
35. Z. Hao, E. Novak, S. Yi, Q. Li, IEEE Internet Computing **21**(2), 44 (2017). https://doi.org/10.1109/MIC.2017.26
36. J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, W. Zhao, IEEE Internet of Things Journal **4**(5), 1125 (2017). https://doi.org/10.1109/JIOT.2017.2683200
37. IDC FutureScape: Worldwide internet of things 2016 predictions. https://www.idc.com/research/viewtoc.jsp?containerId=259856
38. M. Satyanarayanan, P. Bahl, R. Caceres, N. Davies, IEEE Pervasive Computing **8**(4), 14 (2009). https://doi.org/10.1109/MPRV.2009.82
39. G. Paschos, E. Bastug, I. Land, G. Caire, M. Debbah, IEEE Communications Magazine **54**(8), 16 (2016). https://doi.org/10.1109/MCOM.2016.7537172
40. D. Costenaro, D. Duer, in *17th biennial ACEEE conference) on Energy Efficiency in Buildings, 12–17 August, 65–76, Pacific Grove, CA, USA* (2012)
41. F. Jalali, K. Hinton, R. Ayre, T. Alpcan, R.S. Tucker, IEEE Journal on Selected Areas in Communications **34**(5), 1728 (2016). https://doi.org/10.1109/JSAC.2016.2545559
42. M.P. Papazoglou, in *Proceedings of the Fourth International Conference on Web Information Systems Engineering, 2003. WISE 2003.* (2003), pp. 3–12. https://doi.org/10.1109/WISE.2003.1254461
43. W.S. McCulloch, W. Pitts, The Bulletin of Mathematical Biophysics **5**(4), 115 (1943). https://doi.org/10.1007/bf02478259
44. Y. LeCun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard, L.D. Jackel, in *Advances in Neural Information Processing Systems 2*, ed. by D.S. Touretzky (Morgan-Kaufmann, 1990), pp. 396–404. http://papers.nips.cc/paper/ 293-handwritten-digit-recognition-with-a-back-propagation-network.pdf
45. A. Krizhevsky, I. Sutskever, G.E. Hinton, in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1* (Curran Associates Inc., USA, 2012), NIPS'12, pp. 1097–1105. https://dl.acm.org/citation.cfm?id=2999134.2999257
46. G.E. Hinton, S. Osindero, Y.W. Teh, Neural Comput. **18**(7), 1527 (2006). https://doi.org//10.1162/neco.2006.18.7.1527
47. G.E. Hinton, Science **313**(5786), 504 (2006). https://doi.org/10.1126/science.1127647
48. Z.C. Lipton, CoRR **abs/1506.00019** (2015). http://arxiv.org/abs/1506.00019
49. J. Schmidhuber, CoRR **abs/1404.7828** (2014). http://arxiv.org/abs/1404.7828
50. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Journal of Machine Learning Research **15**, 1929 (2014). http://jmlr.org/papers/v15/srivastava14a.html
51. A.Y. Ng, in *Proceedings of the Twenty-first International Conference on Machine Learning* (ACM, New York, NY, USA, 2004), ICML '04, pp. 78–. https://doi.acm.org/10.1145/1015330.1015435
52. M. Riedmiller, H. Braun, in *IEEE International Conference on Neural Networks* (1993), pp. 586–591 vol.1. https://doi.org/10.1109/ICNN.1993.298623
53. J. Duchi, E. Hazan, Y. Singer, J. Mach. Learn. Res. **12**, 2121 (2011). http://dl.acm.org/citation.cfm?id=1953048.2021068
54. S. Ioffe, C. Szegedy, CoRR **abs/1502.03167** (2015). http://arxiv.org/abs/1502.03167
55. J.L. Ba, J.R. Kiros, G.E. Hinton, arXiv preprint (2016). http://arxiv.org/pdf/1607.06450.pdf
56. A.C.C. Yao, in *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)* (1985), pp. 1–10. https://doi.org/10.1109/SFCS.1985.49

57. H. Zhang, Z. Zheng, S. Xu, W. Dai, Q. Ho, X. Liang, Z. Hu, J. Wei, P. Xie, E.P. Xing, in *2017* {USENIX} *Annual Technical Conference ({USENIX} {ATC} 17)* (2017), pp. 181–193
58. S.X. Zou, C.Y. Chen, J.L. Wu, C.N. Chou, C.C. Tsao, K.C. Tung, T.W. Lin, C.L. Sung, E.Y. Chang, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2017)
59. Y. Liu, V. Vlassov, in *2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (2013), pp. 225–232. https://doi.org/10.1109/CyberC.2013.44
60. R. Hasan, Z. Anwar, W. Yurcik, L. Brumbaugh, R. Campbell, in *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II - Volume 02* (IEEE Computer Society, Washington, DC, USA, 2005), ITCC '05, pp. 205–213. https://doi.org/10.1109/ITCC.2005.42
61. D.P. Kingma, S. Mohamed, D. Jimenez Rezende, M. Welling, in *Advances in Neural Information Processing Systems 27*, ed. by Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Curran Associates, Inc., 2014), pp. 3581–3589. http://papers.nips.cc/paper/5352-semi-supervised-learning-with-deep-generative-models.pdf
62. A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, T.P. Lillicrap, CoRR **abs/1605.06065** (2016). http://arxiv.org/abs/1605.06065
63. A. Rasmus, H. Valpola, M. Honkala, M. Berglund, T. Raiko, in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2* (MIT Press, Cambridge, MA, USA, 2015), NIPS'15, pp. 3546–3554. https://dl.acm.org/citation.cfm?id=2969442.2969635
64. J. Weston, F. Ratle, R. Collobert, in *Proceedings of the 25th International Conference on Machine Learning* (ACM, New York, NY, USA, 2008), ICML '08, pp. 1168–1175. https://doi.acm.org/10.1145/1390156.1390303
65. A. Abdelgawad, K. Yelamarthi, Wireless Communications and Mobile Computing **2017** (2017)
66. D.J. Patterson, L. Liao, K. Gajos, M. Collier, N. Livic, K. Olson, S. Wang, D. Fox, H. Kautz, *Opportunity Knocks: A System to Provide Cognitive Assistance with Transportation Services* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004), pp. 433–450
67. K. Ha, Z. Chen, W. Hu, W. Richter, P. Pillai, M. Satyanarayanan, in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services* (ACM, 2014), pp. 68–81
68. K. Bilal, O. Khalid, A. Erbad, S.U. Khan, Computer Networks **130**, 94 https://doi.org/10.1016/j.comnet.2017.10.002. (2018). http://www.sciencedirect.com/science/article/pii/S1389128617303778
69. Reactive manifesto. http://reactivemanifesto.org/. Accessed: 2018-01-13
70. E. Bainomugisha, A.L. Carreton, T.v. Cutsem, S. Mostinckx, W.d. Meuter, ACM Comput. Surv. **45**(4), 52:1 (2013). https://doi.acm.org/10.1145/2501654.2501666
71. C.A. Meier, M.C. Fitzgerald, J.M. Smith, Annual review of biomedical engineering **15**, 359 (2013)
72. O.D. Lara, M.A. Labrador, IEEE Communications Surveys Tutorials **15**(3), 1192 (2013). https://doi.org/10.1109/SURV.2012.110112.00192
73. S. Wolfert, L. Ge, C. Verdouw, M.J. Bogaardt, Agricultural Systems **153**, 69 (2017)
74. A. Kamilaris, F. Gao, F.X. Prenafeta-Boldu, M.I. Ali, in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)* (IEEE, 2016), pp. 442–447
75. S. Trilles, J. Torres-Sospedra, Ó. Belmonte, F.J. Zarazaga-Soria, A. González-Pérez, J. Huerta, Sustainable Computing: Informatics and Systems (2019)
76. R. Bramley, in *Managing Wine Quality: Viticulture and Wine Quality* (Elsevier, 2010), pp. 445–480
77. G. Goidànich, *Manuale di patologia vegetale*, vol. 2 (Edagricole, Bologna, 1964)
78. J. Carroll, W. Wilcox, Phytopathology **93**(9), 1137 (2003)
79. D. Molitor, B. Berkelmann-Loehnertz, Crop Protection **30**(12), 1649 (2011)
80. J. Broome, J. English, J. Marois, B. Latorre, J. Aviles, Phytopathology **85**(1), 97 (1995)

Urban Big Data: City Management and Real Estate Markets

Richard Barkham, Sheharyar Bokhari, and Albert Saiz

1 Introduction

We are in a period of extraordinarily rapid technological change. A great many innovations will transform our cities and the ways we work, live, and move around. Electric autonomous vehicles (AV)—available on demand—will eventually move us all around town. Robots and globalization will keep on having impacts on local labor markets. Smart buildings may make our lives a bit better. Yet a very significant transformation does not entail any physical changes, but data collection and a better use of information.

The power of data to improve the living conditions of urban dwellers was illustrated early on by the series of cholera epidemics in London in the mid nineteenth century. While the medical profession of that time was comfortable in the belief that cholera was transmitted by a mysterious "miasma," Doctor John Snow was busy tenaciously collecting data about where each of the dying patients had lived. Using this data, he quickly realized that mortality from cholera was much higher in buildings that were supplied water from a source that was contaminated by fecal material from waste water. Subsequent data collection efforts proved unambiguously that all persons falling ill to cholera had exposure to contaminated

R. Barkham
CBRE, Dallas, TX, USA

S. Bokhari
Redfin, Seattle, WA, USA

A. Saiz (✉)
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: saiz@mit.edu

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_10

water sources. This pointed to the solutions: better pipe insulation for fresh water and more careful disposal of brown waters. And thus, future epidemics were avoided.

Many more similar advances are now possible by the generalization of data collected by sensors, internet-enabled devices, and the computerization of administrative and business records. Some of the applications are still in the health sector: in New York city, asthma patient records are now analyzed to identify which buildings need to be retrofitted to minimize the presence of mold, and dust mites. Yet the use of urban big data can improve our standard of urban life in many other arenas by creating more efficient transportation systems, reducing energy usage, identifying and addressing criminal activity, reducing government waste, better illuminating streets, and creating environments that adapt and respond to human needs in real time.

1.1 What Is Big Data?

Big data is a buzzword that permeates much of the discussion surrounding the use of digital information in business and government. Surprisingly, there is little consensus on a conceptual definition of the term. Hence—before proceeding further—we provide some alternate perspectives on what constitutes big data.

Perhaps the most academic definition of big data comes from describing its characteristics by the three Vs[1] of big data management:

1. *Volume*: Big data is often characterized by their very large size, often in petabytes or more. "Simply put, big data is too big to sit on your hard drive," says Eric Scharnhorst, a Data Scientist at Redfin.[2] Examples include user-data on Facebook, flow of people through transport system, or credit card transactions.
2. *Velocity*: Big data are often generated continuously, in or near real time. For example, twitter posts, cell phone location, weather, and air quality sensors.
3. *Variety*: Big data can be of any type; consisting of numbers, text, photos, videos, audio, and other kinds of data. Its analysis needs to be flexible enough to deal with any type or, increasingly, often, a mix of these.

While this characterization may be useful in some contexts, the phenomena of big data are recent enough that not all data examples will necessarily hold each of these characteristics. Furthermore, not all experts in the use of urban big data regard it in a similar vein. For example, Kent Larson of the MIT Media Lab thinks very broadly "of big data not in terms of lot of data but complex data that can reveal patterns useful to understand, that aren't evident without analyzing the data."[3] On the other

[1] Kitchen (2014).

[2] Conversation with author, Oct 17th, 2016.

[3] Conversation with author, Sept 28th, 2016.

hand, Jim Sempere of Real Capital Analytics –one of the main data providers in the US real estate industry—emphasizes "... big data as variety of data, put together from various sources to come together in a meaningful way."[4] And Fabio Durate, of the MIT's Senseable City Lab, provides yet another perspective by regarding big data as "... any human activity that leaves digital traces, which does not have an immediate meaning, but when combined together gives some meaningful pattern."[5]

While scale, variety, and frequency are oftentimes necessary for good applications, the three experts' views above do coincide in two critical main characteristics: big data requires *collection* and *analytical efforts*, and must be *socially useful*.

1.2 City Data and Urban Efficiency: Economic Impacts

Urban big data is still in its early stages but what will be clear from the investigation in this report, is that there is already a consolidated global movement around this phenomenon. Oftentimes, observers use the term "smart city" to encompass the places and initiatives that embrace the use of computerized information and analytics to improve municipal and building operations. Of course, it will take a while for urban data systems to be widely adopted in all their potential applications. Nonetheless we are certain to eventually see urban analytics play an important role in many metropolises around the world.

The extent and shape of the implementation of smart city initiatives will have impacts on land use patterns, real estate markets, and the quality of life. It is therefore imperative for practitioners in the real estate industry to anticipate their effects, in order to take advantage of new opportunities, or simply to take corrective action to avoid disruption of established business practices. There are many angles to such "horizon scanning." One of them focuses on the relative advantages of cities that adopt smart city technologies compared to other areas that do not.

Economists have long formulated research on cities in terms of spatial equilibrium. A city is in spatial equilibrium at the point where there is no pay-off for people to move in or out. The vast majority are doing fine where they are. Those who are not —and who might want to move on—would not be better off by leaving either, due to a lack of better opportunities elsewhere. Of course, a highly mobile society always has people hopping across locales. In equilibrium, most of these moves cancel out so that net migration flows are smaller than gross movements.

Where mobility does take place in thriving societies it tends to follow opportunity in ways that are predictable. A wealth of evidence demonstrates that population growth tends to follow wage growth and —increasingly—the availability of amenities. But if the more successful cities tend to attract people: why don't we see most population clustering in just a few metropolitan areas?

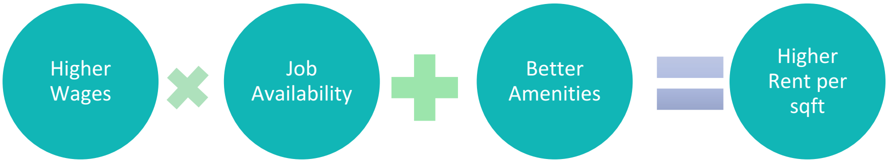
---

4 Conversation with author, Nov 14th, 2016.

5 Conversation with author, Sept 30th, 2016.

The answer partially resides in the role of the housing market. As cities grow larger, they also tend to get more expensive. When a city is somewhere close to its equilibrium, housing price inflation will exactly compensate for the gains in productivity which give rise to higher wages, so that there is no general advantage for even more people to move into more attractive places such as London, New York, Barcelona, or Boston.

In practical terms, this means that if wages and the quality of life in urban areas are much higher than wages in non-urban areas, then we expect higher urban rents for real estate as offsetting this urban wage premium. Glaeser, Kolko, and Saiz[6] formulate a simple equation to capture this concept of a spatial equilibrium:



Productivity Premium + Amenity Premium = Rent Premium.

This equation simply states that the effect of local wages and availability of good jobs (productivity premium) plus the local effect of quality of life (amenity premium) must be offset by local real estate values or rents (rent premium).

This equation is not simply a theoretical identity, as it has been validated in many cities over many periods. It provides a good framework for forecasting the evolution of local real estate markets and for exploring the impact of current trends, such as the application of big data techniques to the management of cities. As an example, consider a case where city wages are growing but in which housing costs are booming even faster. We can infer that the quality of life in the city has risen faster than the rise in productivity. This is exactly the case in a number of global gateway cities for at least the last 20 years. Productivity and wages have grown, but amenities have also grown, possibly faster. As a result the global gateways have become very expensive places to live.

The productivity premium, which drives wages, comes mainly from two sources.[7] One of them is the existence of agglomeration economies. These are efficiency gains in production that arise from having firms be close to each other. More saliently, the clustering of firms by sector allows for flexible supply chains and reduces transportation costs of inputs in industrial production. In the services

[6] Glaeser et al. (2001).

[7] Ibid.

sector, physical proximity facilitates face-to-face communication of high-level information and allows parties involved in complex transactions to develop trust. Parallel to physical input–output relationships between industrial plants, proximity between services firms allows for the presence of specialist sub-contractors and for fast reaction times from service providers. Agglomeration economies also arise with respect to attracting talent: if the best firms in a sector co-locate in a city, that city becomes a beacon for highly-skilled workers, which in turns makes it yet more productive for more firms to locate there. Similar arguments can be made about the access to support networks for new firms, such as computer coding workshops or meet-and-greet cocktail parties.

In parallel, firms can also be more productive in places where they have better access to ideas, technology, and skilled workers. These ideas range from changes in industrial techniques or product types, to the most up-to-the-minute news about prices. This information is often transmitted by workers as they change firms and in social settings like after-work happy hours.

The amenity premium can arise from the provision of four types of amenities.[8] First, an important amenity is the presence of a rich variety of services and consumer goods. A second amenity is aesthetics: physical attributes of a community that make life more pleasant appear to be increasingly valued by consumers. A third amenity is good public services, which include good schools, better public utilities and less crime, and cultural opportunities. The fourth amenity is the speed or ease with which individuals can move around in a city. As people become richer they value their time more, in the sense that they are willing to pay higher amounts for more expeditious services, or that they are willing to pay higher rents to live in locations closer to their jobs.

The framework outlined above suggests that where big data efforts enable greater productivity gains and/or help cities provide better amenities, they will be associated with local real estate value appreciation. Urban data analytics certainly have the potential to improve the provision of urban amenities. Alternatively they could maintain current municipal service levels at relatively low cost. If new IT systems and methods are successful in this latter regard, they could stem the secular "cost disease" of public services: their productivity has historically grown more slowly than in other sectors, rendering them relatively more expensive.

Urban information systems could also help cities organize urban functions so that negative externalities are reduced. Negative externalities are instances where the behavior of some individuals has detrimental impacts on the welfare of others. Negative externalities destroy amenities locally and at the city level. For instance, by using sensors to monitor noise levels in the street and software to contact municipal police whenever some pre-established thresholds are crossed, such nuisances could be swiftly remediated. Similarly, positive externalities—where individual behavior has positive spillovers onto others—can be encouraged. For instance, crowdsourcing platforms can allow citizens to report problems—as in the app in Boston allowing

[8] Ibid.

citizens to report the geolocation of potholes—thereby providing free monitoring services to all neighbors.

Finally, the application of big data may increase the productivity of private firms in the city and generate new business opportunities for them. For instance, when app developers and corporate IT departments gain access to large quantities of data they can develop new software, and build better customer predictive analytics platforms.

This discussion suggests that, either by providing productive value or by improving the local quality of life, the use of big data has the potential to increase urban residential and office demand, and therefore support local real estate valuations in the cities in which it is deployed. Yet, some substantive questions remain:

1. How much value? While urban IT systems can improve our quality of life and reduce economic costs, the real issue is by how much. An additional matter pertains to housing supply. In some markets, an increase in urban demand is easily accommodated by local markets via new real estate development. In these markets, real estate is essentially commoditized and one would not expect pricing to deviate much from construction costs plus an adequate markup to account for development risk. In markets with high barriers to entry, improvements in the attractiveness of the city conversely translate quickly into more expensive land values.
2. Where? Real estate value emanates from differences in productivity across locales. Technological change that has a broad impact everywhere—e.g., the smartphone—will not be captured by differential growth in property prices. It is therefore important to forecast how much heterogeneity in the adoption of big data strategies across cities we will see. Will there be winners and losers? How fast will innovations spread around? Will the implementation of such technologies allow cities that now lag in public services to leapfrog more advanced ones? Conversely, will most advanced cities be faster in adopting the techniques, therefore widening inequalities?
3. A last important issue regarding smart city technologies revolves around their impact within metro areas. Will such technologies have re-centralizing effects, making core central cities more competitive? For instance, motion sensors may be cost-effective in denser urban settings but not in suburban roads. Alternatively, will such technologies favor leapfrogging in the provision of urban services by suburbs or edge cities? Even within central cities, technologies could have a disparate effect by allowing neighborhoods with social problems to catch up on the provision of services, such as safety and beautification.

In this report, we describe some of the new projects and technologies, and briefly explore the ultimate impact of urban IT systems in real estate markets. The rest of the chapter is organized as follows. Section 2 , we focus on the details of ongoing projects, the sampling of which emphasizes the variety of big data efforts that are taking shape. In Sect. 3, we discuss some of the common sources of urban big data and several examples of their use in cities. In Sect. 4, we turn to the discussion of the challenges and issues facing cities in implementing big data strategies and point out some areas that are seeing improvement. Finally, In Sect. 5, we discuss, we discuss the implications of big data use on the future of cities and real estate and what areas of effort might be specifically beneficial for the real estate industry.

2 Big Data and City Management: The Present

This section provides some detail on specific projects or phenomena pertaining to big data that are presently underway around the world. We focus on examples that may directly or indirectly increase the productivity or efficiency of actors in a city. We also highlight cases where a city's policy explicitly considers "place making," an effort where people collectively reimagine and reinvent public spaces. Public investments in leisure spaces and beautification, either through improved amenities or removal of bad amenities such as blight, can spur demographic change, bring economic development, and result in long-run appreciation (or prevent long-run decline) of real estate values. Table 1 below gives several quick examples of initiatives, their improvements to the city, and the potential effect on real estate values in the long run. We then turn to a few case studies, to understand the perspective of the actors involved in the effort.

The examples in Table 1 and later in the chapter illustrate the diversity of such endeavors and the creative processes that big data is unleashing on municipal government management and operations. These examples fall into a number of categories. Some cities are focusing on issues related to private and public health: by measuring and identifying morbidity patterns they are better able to provide preventive care and interventions (Cincinnati), mitigate lead issues and rodent infestations (Chicago), facilitate one-point-access health services (New York, Yinchuan), and improve behavioral health (Providence).

Other efforts focus on environmental concerns and the reduction of waste. Examples include reducing water waste and costs (Santa Clara), mitigating the impact of runoff water on streets and neighbors (Philadelphia), or reducing electricity costs from lighting by rationalizing its use (Buenos Aires, Kansas City).

In another set of applications, budding big data is deployed to deal with issues of public safety and risk prevention, including improvements in the perception of these phenomena by citizens. Monitoring activity of streets (Rio, Songdo), or using artificial intelligence to forecast and thwart future criminal activity (Seattle).

Applications in transportation have also emerged, be it by using big data to facilitate planning (London), providing consumers with information about how to get around public transportation (Nairobi), monitoring for recurring traffic problems in order to mitigate them (Songdo, Andorra).

Land use, property taxes, and registration also occupy a niche in smart city applications: from detection of vacant properties (Detroit), enforcement of codes (Washington, DC), or fair local property taxation (Spain).

2.1 *City Innovation-Teams (I-Teams): Boston Mayor's Office of New Urban Mechanics (MONUM)*

City leaders have begun to recognize the need for better ways to generate ideas. All over the world, innovation-teams (or i-teams) are being created to catalyze innovation. The Boston Mayor's Office of New Urban Mechanics (MONUM) was among the earliest of its kind, launching in 2010. "By now, we are no longer unique.

Table 1 Examples of big data use and their impact on real estate markets

Example of Cities	Impact on City & Real Estate Values
NEW YORK CITY, NY: The Health and Human Services Connect initiative collects all social service data in one place, allowing clients to walk into different agencies duplication of paperwork. The city saves money, provides better services, and better detects fraud.	Improved efficiency of public services → Potential lower property taxes
SANTA CLARA, CA: The drought in California prompted the city to retrofit the municipal irrigation system with sensors in order to more efficiently manage water supplies. The system is expected to save 180 million gallons of water.	Remove negative amenities and save money → Avoid long-term decline in values

(continued)

Table 1 (continued)

CHICAGO, IL: Data Science for Social Good at the University of Chicago runs several initiatives with the city, including predictive analytics for pointing out problems before they happen, such as identifying houses most at risk for lead contamination.	Improved health and safety measures	Avoid long-term decline in values
LONDON, UK: Transport for London (TFL) uses ticketing data to build travel patterns across its rail and bus networks, helping in improving the network and reviewing the impact of closures and diversions.	Better transit flow	May raise values in city peripheral
CINCINNATI, OH: The city has a infant mortality rate (IMR) and is tackling the problem with a data-driven approach to identify the concentration of the problem and explore its causes. Initial efforts have seen a modest reduction in the IMR.	Improved health and safety	Avoid long-term decline in values

(continued)

Table 1 (continued)

Example of Cities	Impact on City & Real Estate Markets	
PROVIDENCE, RI: The city provides low-income families with devices that measure the auditory environment of children and nudges parents/caregivers to increase the number of words the children are hearing.	Improved education	Long-term positive support of residential market
PHILADELPHIA, PA: Lacking a physical divide between storm water and the sewer system, the city implemented rain barrels and street trees to absorb rainfall. A profusion of sensors spread within the city's sewer system provide the data to see if the approach is working.	Remove negative amenities	Avoid long-term decline in values
RIO DE JANEIRO, BRAZIL: IBM has designed the Operations Center of the city that integrates data from 30 different agencies. These provide public services related to safety as well as early warning and evacuation system for Rio's favelas.	Improved public services and safety net	Long-term positive impact on poor neighborhoods

(continued)

Table 1 (continued)

SEATTLE, WA: The city implemented predictive policing software that uses historical data to predict potential crime locations. By focusing patrols in the high-risk areas, the software is expected to limit the need for additional officers on patrol and reduce the number of arrests.	Less crime	Avoid long-term decline in values
DETROIT, MI: The city has very high vacancy rates. Its Blight Removal Task Force surveyed the city's properties, sending information wirelessly to an operations center. This has helped start a data-driven approach to identifying candidates for demolition and blight removal.	Remove negative amenities at lower cost	Avoid long-term decline in values

(continued)

Table 1 (continued)

Example of Cities	Impact on City & Real Estate Markets
WASHINGTON, DC: The Urban Forest Administration used light detection and ranging (lidar) data to identify illegal tree removal based on the original height of the trees. This improved enforcement of permitting laws. **BUENOS AIRES, ARGENTINA**: Philips' CityTouch street lighting management system allows operators to monitor, manage, and analyze information about luminaire performance. It is expected to reduce operational costs by more than 50%. **YINCHUAN, CHINA** : Yinchuan is a smart city pilot project. Features such as facial recognition on buses, grocery delivery via apps, and an online portal connecting doctors with patients have been set up. There is also a command center that oversees a range of data.	City beautification → Better neighborhood quality Improved efficiency of public services → Potential for lower local taxation Increased productivity in city → Make city more attractive and spur residential demand

(continued)

Table 1 (continued)

KANSAS CITY, MO: Cisco, Sprint and Think Big Partners have launched KC Living Lab. The Lab is developing Internet of Things technologies such as smart streetlights, smart traffic signals and interactive kiosks.	
SPAIN: The Spanish tax agency analyzed imagery from unmanned drones across 4,000 municipalities to discover 1.69m properties paying insufficient taxes on new constructions, expansions and pools. The initiative brought in 1.2bn euros in additional taxes.	

Pretty much every major city in the United States has some version of what we do,"[9] says Director Nigel Jacobs. MONUM was initially created by public funding but since then has received significant private support as well. It serves as a good model for what cities are striving to do. Nigel Jacobs speaks of their effort as "we are always trying to track as closely as we can to the important items; the big, strategic issues." The team is given the freedom to be experimental, to collaborate with third parties and is structured to be flexible enough to adapt to new developments.

Recently, MONUM's work was divided across three major initiatives:

- *Platform for City-Owned Property*: This initiative will create a central platform with detailed information on every city-owned property (including vacant lots)

[9] Conversation with author, Nov second, 2016.

belonging to various stakeholders and identify areas of opportunity. For example, an under-used parking lot residing within the transport department may better meet the requirements of the neighborhood development team looking for opportunities to build affordable housing. "We are trying to both start a conversation and also provide a tool that would allow the directors and planners of the various departments to visualize where the opportunities lie that align with the priorities of the mayor," explains Ilona Kramer, Program Director.[10] The platform will not only collect information on the properties but will help developers to identify where the highest value lies and gather the programmatic goals of the various city departments. Ilona believes that there is also potential in creating the highest value in places where third parties are less interested; "The Bolling building in Roxbury is a very good example, the new location for the headquarters of the Boston Public School system. It leveraged state and federal tax financing incentives and required many different departments to work together in areas a developer wouldn't necessarily be interested in, but are viable because of the available policy incentives." This new platform will enable the city to mix and match objectives with policy incentives (and to propose new ones) in order to create impactful outcomes.

- *Sense Lab*: The goal of this lab is to create a network of sensor arrays to better understand public spaces. One test bed is a network of eleven sensors at Downtown Crossing, a popular shopping district with a dedicated pedestrian zone. The data from the sensors will be used to explore ways to make the area more pedestrian friendly. The sensors take people counts, activity spots, environment quality, and luminosity throughout the day. The lab is also correlating their data with various kinds of socio-economic records in order to study larger design questions. "We hope to make a portable array of sensors that we can bring to neighborhoods around the city,"[11] says Stephen Walter, Program Director, in speaking about making sensors more ubiquitous in the planning process.
- *Third Spaces Lab*: Third places are everything outside of the home and work, particularly public and social spaces where people interact, such as public parks, coffee shops, and even community gatherings in private spaces. The main goal of the third spaces lab is to improve the quality of life of a neighborhood. It is exploring how government can support third spaces and aims to use its funding to create design prototypes to make these spaces more vibrant, equitable, and resilient over time. Program Director Susan Nyugen emphasizes that their initial focus is not on showing an economic benefit of investments in third spaces but rather on improving their quality of life, "We will be coupling the quantitative investment analysis of the space with the qualitative experiences of the space. We want to figure out how a quantitative investment can actually turn into a qualitative effect."[12]

[10] Conversation with author, Nov second, 2016.

[11] Conversation with author, Nov second, 2016.

[12] Conversation with author, Nov second, 2016.

While these are the major current projects underway at MONUM, there have been several impactful projects in the past. We will return to discuss some of them further in later chapters.

2.2 Planned "Smart" Cities: Korea's Songdo International Business District

Following Anthony Townsend's broad definition, smart cities are " . . . places where information technology is combined with infrastructure, architecture, everyday objects, and even our bodies to address social, economic, and environmental problems."[13] Some major examples of master-planned smart cities are Konza Technology City in Kenya, Forrest City in Malaysia, Masdar City in Abu Dhabi, and Songdo City in Korea. In addition, both China and India have announced plans to build several smart cities. There are nearly 200 pilot projects of smart cities in China[14] and the Indian government has announced plans to develop 100 smart cities.[15]

After 15 years of development since 2001, Korea's Songdo International Business District (IBD) housed more than 45,000 residents and over 1500 businesses.[16] The project has been developed by Gale International with the support of the Incheon municipal government and is expected to be fully completed by 2018. The new city is part of the larger Incheon Free Economic Zone and is connected to the Korean mainland by bridges. It is strategically located close to the Incheon International Airport, which boasts 90 min flights to Shanghai and Tokyo. According to Andrew Snowhite, Senior Vice President at Gale International, the master-developer designed and built over 65 million square feet of buildings in the International Business District, with roughly 22 million square feet in Leadership in Energy and Environmental Design (LEED) certified buildings. In total, 40% of the land is devoted to green spaces, including a 101-acre central park with a self-sustaining irrigation system.

Mr. Snowhite explains that –initially—Incheon City built the infrastructure around Songdo and created a company called Incheon U-City Corporation to provide the technology overlay. It laid fiber optics in the ground and systems monitoring services throughout the city. A myriad of sensors has been placed throughout the city, from which data on environmental pollution, safety, and security are analyzed and monitored from an operations center. The data collected allows

[13] Townsend (2013).

[14] Yongling et al. (2015).

[15] Government of India (2015)

[16] Gale International's Press Release (2016) http://www.prnewswire.com/news-releases/songdo-international-business-district-to-be-featured-at-greenbuild-2016-as-exemplar-of-sustainable-new-city-300338912.html

for automatized service delivery (e.g., computers automatically calling the fire department upon sensors detecting fire).

The bridges connected to mainland Korea host cameras that collect photos of license plates of every vehicle entering Songdo, allowing for automatic tolling, better traffic control, and security. Surveillance systems throughout the city have face recognition capabilities. While the potential for misuse of this technology is obvious, it could also be used to tailor specific services to each citizen (e.g., by opening specific doors on sight). We discuss the privacy concerns surrounding these measures in Sect. 4.2. The city also has the world's largest pneumatic waste disposal system, which bypasses the need for garbage collection. In order to create ubiquitous technology to improve people's lives, Gale International, in collaboration with Cisco, created a company called U.Life Solutions to install building management systems, with security sensors, greywater facility, and lighting sensors. "Each residential apartment has hardware on the wall that manages energy usage in order to reduce costs, electronics in the house, provides a closed circuit television (CCTV) channel with a view of the playground (to monitor children's safety), and a Telepresence service that you can opt into to take classes (with a personal tutor) for sewing or learning English," says Andrew Snowhite.[17] Gale International estimates that the overall energy in Songdo IBD is up to 40% less per person than an average existing city due to better insulation, high-performance glass, and high-tech equipment for lighting, heating, and air-conditioning.

It is unclear that all the technological solutions in Songdo are cost-effective and broadly generalizable, or even desired by wide segments of the population. They certainly provide a maximalist example of what a connected city can be nowadays, and a real estate product that can appeal to the sensitivity of Korean consumers. Songdo can also be seen as a pioneering effort from which other smart cities will no doubt learn.

2.3 Analyzing People Behavior: Andorra

Andorra is a small country between France and Spain with a tourism-focused economy. It is also the location of a unique collaboration between the MIT Media Lab, the Andorra's government and international companies such as General Electric. The goal of the collaboration is to apply data science methodologies in understanding the country's dynamics on tourism, commerce, human mobility, transportation systems, and environmental impact; as well as to look for radical improvements in these areas. Kent Larson of the MIT Media Lab contrasts the approach of his team in Andorra with that of smart cities as "I think of it (big data) as two ends of a spectrum. One is the typical smart city's use of it, which is more about optimization; improving the throughput of cars on the roadways and what not.

[17] Conversation with author, Oct 11th, 2016.

And then there is the use of data to gain new insights on how the city functions and how to improve the fundamentals of the city.[18]" Towards that effort, he describes the following projects underway:

- **Tourist Recommendation System**: This project integrates data from cell phones, social media, bank transactions, energy consumption, and transportation to study the flow and behavior of people. The analysis generated from this big data platform is used to predict tourist behavior and routes in Andorra to develop a targeted tourist attraction recommendation system.
- **Traffic Flow**: Since Andorra has no airport or train service, most of the eight million tourists visiting the country arrive by car, making traffic management and car parking important challenges. Having cell phone data of the tourists and equipped with the knowledge of their mobility (how they move around the city, how long they stay, where do they come from, etc.) and transportation data, a traffic prediction tool is being created to avoid congestion. There is also a pilot program for deploying autonomous vehicles for transporting tourists.[19]

One of the end-goals of the initiative is for Andorran authorities to have the capability to effectively use these data on their own. Later in Sect. 4, we discuss human capital constraints on cities, so it is important to underscore that this is a unique example of a participatory collaboration where expertise is handed over after the initial phase of development is over.

2.4 Transportation: Nairobi, Kenya

In many developing countries, the public transit system is "informal," with basic information about schedules, routes, and fares not readily available. The Kenyan capital of Nairobi has over 3.5 million people and many rely on the "matatus"—public service vehicles such as minibusses—to get around. These are owned by many private and poorly regulated businesses. Schedules and ticket prices can change without any notice and a lot of insider knowledge is required to navigate the network in an efficient way. In an effort to address these shortcomings, a collaboration between the University of Nairobi, Columbia University, and MIT's Civic Data Design Lab, launched a project called Digital Matatus (or the digitization of the public transit system).

To collect data, Kenyan students rode over 130 Matatus routes with smartphones in hand and a GIS unit that recorded data every couple of seconds.[20] They carefully geo-located stops and collected key information about the route. "The data developed in this project was launched in Google Maps, making it the first

[18] Conversation with author, Sept 28th, 2016.

[19] Ara andoraa (2016)

[20] Willams et al. (2015).

informal transit in which Google has provided transit routing directions," says Sarah Williams of MIT's Civic Data Design Lab.[21] Professor Williams and her colleagues also converted the data into a subway-style map that shows all the routes and main stops. The paper map and the transit data were released for free and have found use by riders, Kenyan officials, and app developers.

While it was virtually impossible for many users to figure out combinations of 2 or 3 rides to reach to every point of the city, they can now do so. Many African cities consist of very expensive downtown cores, which are unaffordable to the middle class and migrants, and large –often informal—suburbs that extend for miles. By facilitating mobility and access to jobs, travel digital information technologies improve the lives of suburban dwellers. Improvements in their earning capacity can eventually allow them to build better homes in suburban locations and raise neighborhood values. At the same time, increased accessibility and broader options may mute the affordability issues downtown. Thus, smart cities that have a positive impact on suburban transportation may contribute to a "flattening" of the real estate value curve within the city. Of course, if the improvements are large enough, this will be accompanied by even more migration into the city and a general increase in land values in mega-cities.

2.5 Real Estate Technology: BuildZoom

Irrespective of the city, big data can help improve the efficiency of the real estate sector itself. Buildzoom is a real estate technology company that operates in the so-called home services market, remodeling jobs that typically require city permits and experienced contractors. Buildzoom makes use of big data analytics to match commercial or residential owner projects with appropriate contractors: those who specialize in the job at hand, are geographically close, and have high ratings from satisfied customers. The company is able to do that by collecting data from thousands of government sources to create comprehensive profiles for every single licensed contractor in the USA. It is estimated that there are over 110 million building permits, over a million customer recommendations and 3.5 million state license records. Buildzoom is part of a much larger technology trend where big data analytics is being used to bring more objectivity and transparency to decision-making, enhancing the matching process in the market for local contractors.

Buildzoom also uses its data to provide intelligence about the pace of remodeling work in each major city of the USA. For instance, a large contractor or do-it-yourself firm can now learn about the number of kitchen renovations in each city of the Chicago metropolitan area in the last month, thereby improving their marketing and provisioning strategies.

[21] Conversation with author, Sept 27th, 2016.

3 Big Data and City Management: Sources of Data

The rise of big data has been enabled by the meeting of a number of key trends: computation power has grown exponentially relative to cost; computers and digital devices are increasingly networked together, making it easy to transfer and share data; many new technologies such as sensors and swipe cards have made data machine-readable so each datum is uniquely identifiable and easily interconnected with other databases; computer algorithms have exponentially improved with data and computer power; and finally, data storage has expanded, becoming considerably cheaper over time.

Thus, a big data platform may store data from various sources such as sensors, apps, geographic information systems (GIS), images and videos, user behavior (e.g., social media) or crowdsourcing efforts, and make them indexical so that different pieces of information can be queried together. The data collection may be in real time or at irregular intervals (such as monthly), with typical databases growing by several gigabytes or terabytes per frequency of collection.

In this section, we illustrate the recent proliferation of urban data with examples from different sources, and suggest their possible impact on enabling improvements in the city and its real estate values. These examples are not meant to be exhaustive and we can expect that the creativity with which people tackle local problems will give rise to new ways of data generation in the future.

3.1 Administrative

In an effort to improve transparency and accountability, several cities are moving towards a culture of open data. A common effort across several cities is the digitization and compilation of existing administrative data that may be spread across various city departments. For e.g., in New York City, the open-data platform has an identifiable code for each building and each parcel. Each of these can in turn be linked with real estate transactions, leases, capital expenditures along with a plethora of city related information such as data from every city department, air pollution measurements from sensors, taxi and limousine trips from the buildings, public work orders, and even the type of trees on the sidewalk to name a few. These data can be used in a myriad of ways from app development, to monitoring of pollution and congestion to better identifying inefficiency in public provision. OpenGov, a startup from Palo Alto is now working with over 100 local governments to develop city-manager dashboards incorporating real-time data from municipal operations and finances. Their open-data platform allows each citizen to access the information.

Furthermore, many cities are moving away from paper processes to online ones. The City of Cambridge in Massachusetts now has an online voting platform for its residents to democratically allocate $700,000 on capital projects to improve the community. During the idea collection phase, ideas can be submitted via an interactive online map. The data is processed and the final list of projects for voting is once again shown on the interactive map for selection during the voting period. Such easy to use open-data portals that focus on citizen engagement, transparency, and accountability have the potential of saving cities money through more efficient operations, better communication, more competitive contracting, and lower risk of fraud or abuse.

3.2 Sensors

Sensors are perhaps becoming the most well-known form of collecting urban big data. They are placed in structures and buildings to measure a variety of output, ranging from light, temperature, and air quality, to movement and speed of people in the area. Bitsence, a startup company has installed environmental sensors at Downtown Crossing in Boston to measure dust, oxygen and carbon dioxide levels, noise pollution, and even count the number of people at a given distance from the sensor at any given time. Ariana Salazar, a co-founder of Bitsence, says the city has allowed them their platform to come up with innovative ways to use this data. "Our partnership with the city is actually trying to solve the problem of providing data analytics."[22] Bitsence will create a platform from which this data can be analyzed.

Yet another firm partnering with cities is Phillips Lighting. Talmai Oliveria, a Senior Researcher at Phillips Research, says "light is ubiquitous and one of our goals is to leverage light to make lives better, to make streets safer and make light improvements to bring people together."[23] In Cambridge, Massachusetts, Phillip's light and sound sensors at intersections are currently being tested to identify which ones are at risk. The urban flow measured at these intersections will be fed into a dashboard that will control and experiment with directed lighting to understand how to help drivers see pedestrians and bikers more easily. Lighting configurations helping traffic and pedestrian flows improve and can then be widely adopted.

In New York City, fiber optic sensors monitor the cracks on The Brooklyn Bridge as well as other indicators such as temperature fluctuation. This information helps structural engineers in determining when the vaults will ultimately need to be replaced. The Williamsburg bridge also has sensors that provide continuous data to engineers on the century-old suspension cables. These systems help minimize costs associated with unnecessary inspections, and avoid under maintenance, which could lead to very expensive emergency repairs or catastrophic events.

[22] Conversation with author, Oct 14, 2016.

[23] Conversation with author, Oct 21, 2016.

3.3 Apps

Smartphone apps are also emerging as new big data conduits. Software on the IOs and Android platforms is being deployed to both make use of existing information (use data output) or to collect or produce new intelligence (use to input data).

New uses smartphones as tools that use data outputs from online sources are proliferating. With the advent of open transit data, there are now plenty of public transit apps in major cities that provide real time information to riders. Once the data from the Digital Matatus project, discussed in Sect. 2.4, was released in Nairobi, five high-quality mobile apps were immediately launched by the local tech community to serve the general public. Other examples are provided by apps that map local crime rates using real time data provided by local police reports.

Smartphones also allow for users to consciously input data, or for applications to monitor their behavior. Internet connectivity then allows centralized databases to collect such data. In Boston, the Mayor's Office of New Urban Mechanics created a crowdsourcing mobile app called Street Bump that helps residents improve their neighborhood streets by inputting road condition data while they drive. The data is collected and analyzed in real time to provide information to the city on locations that need to be further investigated by employees. Yet another app, Boston311, allows citizens to report city problems from mobile devices. It currently has over 50,000 users. A central system collects all the reports and then funnels them to the appropriate public works department. Several other cities such as San Francisco, Chicago, New York, and Seattle have similar apps.

There is now also a concerted effort to make apps more "civic." In the aftermath of the hurricane Sandy, the Airbnb community came together to provide emergency housing. Airbnb now official runs a Disaster Response Service to make it easier for its members to provide emergency accommodations in times of crisis.[24]

3.4 Crowdsourcing

Crowdsourcing is where a large number of people, often volunteers, produce data collectively on particular issues. One form of crowdsourcing is known as citizen science, where volunteers generate, prepare, and process observations and detailed measurements of some phenomenon. In a study by Gehl Architects,[25] volunteers recorded the activity and flow of people in downtown Denver. A striking conclusion from the study was that the introduction of public art, such as a sculpture on the street, actually reduced public pedestrian traffic. Instead, expanding café patios all the way to the street increased public activity. By pinpointing the best initiatives to

[24] See more at https://www.airbnb.com/disaster-response

[25] Gehl Architects (2015), Downtown Denver 16 h St. Mall: Small Steps Towards Big Change. Report.

activate street life, such big data methods can increase pedestrian traffic and enhance the value of main street retail locations. A vibrant street scene is also a key element in the revitalization and redevelopment of residential neighborhoods.

The Beijing Air Tracks project[26] was a collaboration between the MIT Civic Media Lab and the Associated Press, where journalists attending the Beijing Olympics were armed with handheld aerosol-monitoring devices and a GPS unit. These tracked both air quality and the journalists' geographical location as they moved throughout the city in their daily reporting. The images produced from a month's worth of data were invaluable in understanding the larger patterns of regional pollution. The results provided independent confirmation of the poor air quality, while the local government was largely attributing it as simply fog, not smog. By providing transparency, such applications can pressure municipal governments into action into improving the quality of life of residents.

3.5 Remote Image, GIS, and User Behavior

Given the explosion in the number and variety of digital devices in use around the world it is not surprising that they provide ways of interaction with each other that further create sophisticated datasets. The Atlas of Lighting[27] is a joint project between Philips, MIT Center for Advanced Urbanism, and MIT Civic Data Design Lab that is looking at how lighting intensity varies by various types of socio-economic conditions in the Chicago area. An interactive mapping tool brings together statistical data on demographics, intensity of urban development, and night-time light intensity using satellite imagery. The tool is integrated within Google Places and uses geo-tagged Instagram posts as proxy for human activity. It is also intended to be used as an engagement tool for the public and policy makers in the planning of business districts.

The city of Louisville, KY has a large population suffering from breathing disorders. In order to address the problem, the city distributed 500 smart inhalers to asthmatic residents that recorded the time and place of when those devices were used. Louisville was one of the cities chosen in IBM's smart cities challenge; thus, data analysts from IBM analyzed the data sent by the devices to create "heat maps" of asthma emergency attacks and suggested interventions such as SMART inhalers that let residents track where and when they are having a hard time breathing. The maps also facilitate avoidance behavior. It is estimated that asthma symptoms declined by 43%.[28]

User-generated data from social media such as Twitter and chip readers such as transit cards are now being used extensively in research on people's behavior. As

[26] Williams (2013)

[27] See more at http://civicdatadesignlab.org

[28] Eells and Fletcher (2016)

we learned from the example of Andorra in Sect. 1.3, these sources are increasingly becoming important in analyzing cities. For example, the city of Los Angeles mines texts on its officials' social media accounts to understand what its citizen are most concerned about, what topics are trending in public discourse and how their public relations team should respond or prioritize their focus.

4 Challenges and Issues

It is not all progress towards a world of better information and greater efficiency. The wholescale collection of data, and the use of it, raises profound concerns about the nature of privacy. Not everyone is convinced about the benefits of big data led policies either. This section discusses the range of issues facing the urban big data movement. In particular, we document the opposition to making the data public, privacy concerns, human capital or technology constraints, and the general skepticism about the efficacy of popular big data solutions to urban problems.

4.1 Open Data: Use and Usability

The open-data movement is based on the three principles of transparency, participation, and collaboration.[29] While it is believed that through openness, sharing and working together, the value of data to society can be truly realized, the rapid opening of government data has not been universally welcomed. How open-data projects become funded in the absence of clear revenue streams is a challenge facing many cities. Furthermore, "when resources are limited, data can become a commodity,"[30] says Sarah Williams of MIT's Civic Data Lab, about the unwillingness of city departments to share information. In many cases, taxes pay for the process of integrating and opening the data. In the USA, Socrata, a third party private company is sometimes employed to create data platforms on behalf of municipal governments.

However, there are still many open-data websites that are simply file dumps. These unstructured sites do not follow any standard or even give any attention to the usability of the content. Part of the reason is that they have been built by enthusiasts, such as during hackathons and data dives.[31] In these cases, unfortunately, there is often no post-event follow up, maintenance, or further development. As a result, data use often drops quite markedly after an initial spark of interest.

[29] See White House Directive https://www.whitehouse.gov/open/documents/open-government-directive

[30] Conversation with author, Sept 27th, 2016.

[31] Townsend (2013).

Implicit in the discussion on open data is the notion that data is neutral and that everyone has the potential to access and use it. This is not always the case. Not all citizens have either the skills to process and interpret open-data sets or have the ability to voice their opinions about them. For example, the digitization of land records in Bangalore, India, exemplifies an open-data project that was promoted as a "pro-poor" initiative. However, in some cases it ended up being put to use by upper income people and organizations to gain ownership of land from the poor through taking advantage of mistakes in documents and instructing land surveyors as well as lawyers to challenge titles.[32] In this case, open data may have facilitated a change in land rights from the poor to the rich.

4.2 Privacy

In much of the developed world, privacy is considered a basic human right and protected by law. However, the concept of privacy is in flux. People's behavior is constantly recorded in an increasingly greater number of ways: through social media, the Internet of Things, GPS, transaction records, credit cards, and more. As Talmai Oliveria, Senior Researcher at Phillips puts it, "Privacy is relevant but we are at a transition point."[33]

How privacy is dealt with in big data analysis varies from case to case. "There are always privacy issues .... We have to very carefully de-identify our data, but in a way that allows us to us do what we want. We may want to know that someone from Barcelona bought shoes in Andorra of this amount and at this time, but not exactly the identity of the person or the shoe store," explains MIT's Kent Larson about the complications in using data that is sourced from multiple people and organizations.[34] Andrew Snowhite of Gale International says that the South Korean government has placed strict regulations on the use of data collected from sensors, cameras, and face recognition software in Songdo city. He says, "Data is kept for one month and then deleted."[35] Talmai Oliveria also says that at Phillips there are privacy and security checks on data at every stage, from acquisition and transfer to analytics and storage.

In light of the ubiquitous nature of data collection, cities are beginning to think more broadly about how to lay guidelines around privacy. With regard to sensors in public spaces, Stephen Walter, Program Director at Boston's Mayor's New Urban Mechanics explains: "right now, we have stickers on the street letting people know that there are sensors. But we are also developing a smart cities playbook to create a series of guidelines for everything, from vendors to the public sitting out in the

[32] Gurstein (2011)

[33] Conversation with author, Oct 21st, 2016.

[34] Conversation with author, Sept 28th, 2016.

[35] Conversation with author, Oct 11th, 2016.

streets. What are your rights and privacy in public? What is the social contract there? You are giving something to us, some data, what do you get in return? We are trying to think more clearly about how that looks like and looking into creating a community advisory board that oversees the various sensor projects."[36]

It is still early to fully predict how this new data Commons might shape up.

4.3 Technology and Human Capital Constraints

"Building a large, useful, machine-readable, and meaningful data portal is a non-trivial technical task," writes Brett Goldstein, former Chief Data Officer of Chicago's Department of Innovation and Technology.[37] While companies such as Socrata can provide cities with a ready-to-go platform, Goldstein asserts that cities need to go beyond that to make the platform sustainable without the role of a middleman. "Ninety percent of the data that goes onto data.cityofchicago.org arrives there automatically," says Goldstein. In order to build an effective platform, cities need to be aware of their staff's capabilities and have the available funding to make it happen. As Susan Nyugen of MONUM puts it, "You need to think about what kind of functions do you need on the team to do this effectively?"[38] In an effort to address technological and human constraints, Bloomberg Philanthropies' What Work Cities initiative is developing customized approaches to help select cities expand their use of data.

A particular technological effort that ended in failure was an attempt to build a citywide wireless network in Philadelphia. This is because Wi-Fi was never designed for large-scale, seamless outdoor networks. Partly as a result of that experience and also due to the lack of technical expertise, many cities today seek less risky ways, and a growing number are fishing for existing useful apps that use government data. Apps built around the public transit system are the most successful, making it possible to build viable businesses that leverage open government data. However, many apps built via contests are quickly abandoned. Anthony Townsend writes that the problem with apps is that "they rely on programmers to define problems, instead of citizens or even government itself."[39] Nevertheless, app development contests have catalyzed a community of technologists inside and outside government trying to improve the lives of residents.

[36] Conversation with author, Nov second, 2016.

[37] Goldstein (2013).

[38] Conversation with author, Nov second, 2016.

[39] Townsend (2013).

4.4 Solutions in Search of a Problem

A major trend in urban big data is the presence of private corporations such as IBM, Cisco, Siemens, Intel, Microsoft, and more that have become active players in city management. This is either through being key partners in building new smart cities or partnering with established cities to retrofit their infrastructure with Information and Communications Technology (ICT) and data solutions. While such companies might be fostering innovative and useful interventions, there are concerns about their foray into roles traditionally delivered by the city. “They do not spend enough time understanding the unique problems that a certain city has,”[40] says, Stephen Walter of MONUM. In working with sales teams from various corporations, Susan Nyugen of MONUM finds that “there is a quick jump to conclusion that the solution to any problem and every problem is a technology solution.” She believes that this “surface level understanding of big data is not going to produce any results that we as a team care about or we as a city care about.[41]”

Director Nigel Jacobs of MONUM, who is a computer scientist by training, says “There is also the problem of government people not pushing back on these products. Data visualization is often used to obfuscate or give a very superficial level of understanding. It is literally eye candy, what do you do with that? There is a need for government people to start asking how is that visualization going to connect to people’s lives.”[42]

Talmai Oliveria, the Senior Researcher at Phillips, says in relation to the big data analysis he performs, “We understand that just putting a sensor will not solve any problem. That is why we sift through the data and try to find a value to the end user; find proof that there is an impact on the city.[43]” Mr. Oliveria points out that collaboration with the community is an essential part of this process.

5 The Future: Impact of Big Data and Further Opportunities

In this section, we discuss the places that are leading the front on big data use and what we can expect in terms of its potential impact on different types of cities and on their real estate values. We conclude with a look ahead at some areas where big data collection can contribute to the real estate industry in general.

[40] Conversation with author, Nov second, 2016.

[41] Conversation with author, Nov second, 2016.

[42] Conversation with author, Nov second, 2016.

[43] Conversation with author, Oct 21st, 2016.

5.1 Impact on Cities and Real Estate Values

Currently in the USA, cities such as Boston, New York, New Orleans, Los Angeles, and Chicago are at the forefront of urban big data use. They have successfully overcome the financial, human and technological constraints in setting themselves up to make data-driven decisions. As mentioned in several of the examples in previous sections, they are able to mine user-generated data via social media, apps, card readers, etc. and create integrated data platforms that allow them to better understand the issues facing their communities and as a result, be more efficient and productive in providing public services.

The main implications of the urban information systems revolutions will be an increase in the quality of local public services, and the potential to stem the secular cost disease affecting their provision. We are not quite there yet, because fully unleashing the power of data analytics will require: better coordination across government agencies, standardization of data and software, and moving along the learning curve with regard to which applications are cost effective. Nevertheless, efficient data-driven approaches, sensors, and automated systems will no doubt become more prevalent.

If we recall the framework introduced in the introduction of this report, the implication of big data use is that it could help to adopt cities be more productive and raise their quality of life, thereby potentially increasing local real estate values. However, the value of real estate is local, and relative. Therefore, we only expect large impacts in as much as such innovations are unevenly spread.

There are reasons to believe that this will be so. A disproportionate number of the early adaptors are cities with a combination of forward-looking leadership, multicultural open-minded citizenry, an already active street life, and a high-tech worker base. Therefore, at least for some time, the local implementation of urban information systems to improve quality of life will probably become another force for heterogeneity in city attractiveness, likely reinforcing existing trends. “Consumer cities” high higher concentrations of people in the “creative class”[44] will be likely to be early adopters, making their urban environments even more attractive to the high-skilled workers on which they are coming to depend.

Major super-star cities[45] in which new real estate development encounters severe barriers to entry are likely to become yet-more-expensive 24/7 public service locales: London, Boston, New York, Seattle, Portland, Tokyo, Dubai, San Francisco, Mumbai, Moscow, and others.

However, in a number of smaller creative cities with elastic housing supply, implementing smart city concepts will be feasible without such steep price growth. In places such as Austin (TX), Hyderabad (India), Tallinn (Estonia), Boise (ID), Provo (UT), an available local pool of talents can be tapped to deploy IT and

---

[44] Florida (2002).

[45] Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. “Superstar Cities.” *American Economic Journal: Economic Policy*, 5(4): 167–99.

improve the local quality of life. While real estate valuations in these areas will grow, they may still keep competitive due to increased supply and substantial real estate development.

Cities with relatively low levels of human capital may be behind this trend for some time. Of course, these technologies have the potential for urban laggards – such as some struggling post-industrial cities—to leapfrog in terms of quality of life. However, the instances where this happens may be relatively rare in the medium run, and confined to cities with strong leadership and vision, enlightened local business communities, access to strong institutions of higher learning, or cities receiving strong regional or national subsidies for this purpose.

The potential for leapfrogging is larger in the developing world although the challenges outlined in Sect. 4 are greatly magnified in developing countries, especially when it comes to monetary and technological constraints. Therefore, we will see leapfrogging, and the appearance of a number of leading smart cities in emerging countries, particularly those with the necessary educational, financial, and political assets. China is a likely exception because the urban system is still –to a large extent—impacted by central government decisions. Implementation there will probably come through a top-down process.

We believe that big data collection also has the potential to provide firms with valuable information, business services, and provide opportunities for the development of local apps. The "low-hanging fruit" of urban big data still lies in the innumerable opportunities for the improvement of municipal services. We expect stronger synergies with firms outside of the IT sector to play an ever larger part in urban productivity growth.

The use of big data in smaller scale place making efforts can have an impact on real estate valuations within cities. Most municipal governments will not have the capabilities to move toward full smart city concepts. Nevertheless, easy-to-implement local interventions will abound. As we have seen in some of the examples in this chapter, such interventions tend to focus on improvements on relatively impoverished neighborhoods (health, crime, transportation). Therefore, it is possible that such efforts are associated with rent and price convergence, whereby marginalized neighborhoods be made more attractive.

One last question relates to the impact of urban IT applications on suburbanization and sprawl. Interestingly, many 'smart city' interventions focusing on the quality of life and place making are likely to be more successful in densely-urbanized areas. Sensors, lighting, traffic and pedestrian flows, accurate statistics, crowdsourcing opinions may be too dispersed in the suburbs to justify the cost. These would seem to be a force toward densification and higher residential and retail rents in denser locations.

However, urban IT systems geared towards improvements in transportation systems could have the opposite effect. By lowering transportation costs, they could assist with furthering decentralizing employment. Urban data systems could also improve the functioning of autonomous vehicles.[46] In some of the examples in this chapter, apps (such as carpooling, or those providing information about bus movements) and digital maps have had a positive impact on accessibility to job opportunities from remote locales. By leveling the playfield, these technologies could flatten the real estate value curve, making suburban locations relatively more attractive.

Therefore, the ultimate impact of smart city technologies on the relative growth of suburban versus central real estate values depends on how they affect urban amenities and transportation. Improvements in public service and amenities will tend to have a centralizing effect, whereas reductions in transportation costs should have a centripetal one.

5.2 Examples of Further Opportunities

Several potential sources of big data exist today in forms that require seemingly small incremental investment to become useful to both the public and private sector users. This section discusses three of these and the opportunities that they present. They are meant as examples of a vast set of opportunities that urban big data opens up. We expect an explosion of creativity in the next decade, with hundreds of similar and better ideas emerging.

5.2.1 Leases: The Other Half of Property Transactions

For the last several decades large databases have evolved that track the sales transactions of properties, both residential and commercial. This big data source exists because of title laws and the corresponding requirements for property ownership registration. Concatenating data sources across local registries has contributed greatly to market transparency and efficiency. The same has not been true for lease transactions either residential or commercial. As a consequence, leasing is far less transparent and much less understood than ownership. In large measure this is due to the fact that leases are simply private contracts between tenant and landlord, with no requirements to register or report them.

With greater lease transparency, tenant and landlord could spend less time on rental negotiations, speeding up the process of re-location. New and innovative lease features could be more quickly be assessed, analyzed, and adopted to the benefit of both parties.

[46] Saiz and Salazar (2017).

An interesting issue pertains to the impact of open-source big data on market intermediation. For instance, assume that regulations allow for the appearance of a public repository of leases with standardized fields for each potential contractual contingency. Would this development facilitate parties to advertise and search for space, and sign contracts online thereafter without recourse to real estate brokers? We believe that the answer to this question is no. After all, the proliferation of online information about residential properties in platforms such as Zillow has hardly affected the brokerage model of housing sales.

In fact –beyond their impact on city government—information technologies and big data are also transforming private business models in the real estate sector. But as disparate and incompatible applications proliferate, the need for standardization and consolidation grows. Furthermore, it is often the case that only a few players have the expertise and scale to collect and process data, or to create online exchanges. For instance, the company Costar Group has acquired a dominant market position in the listing of commercial properties for sale and lease via their online platform LoopNet. Albert Saiz and Arianna Salazar[47] predict that "the proliferation of software packages aiding in –or even automatizing parts of– the development, construction, asset management, leasing, zoning, and underwriting services will keep apace. In due time, comprehensive online platforms will consolidate these applications, led by current major real estate service providers and—perhaps—a few entrants from the IT world."

5.2.2 Urban Travel Patterns: What to Do without Census Data?

For more than 40 years, US transportation planning has relied on using the US Census Long Form (a 5% subsample) to identify home-workplace commuting patterns. With the abandonment of the Long Form (and substitution with the far smaller American Community Survey [ACS]) such commuting pattern data will no longer available. Savings on transportation costs in terms of money and time are the main force behind the very need for urbanization. Firms and households make critical location decisions in part based on travel costs. Real estate rents and asset prices in core areas mostly capture a premium to pay in order to avoid high commuting and transportation costs further on. Providing and improving transportation infrastructure not only enables cities to operate better, but also enhances productivity for the country as a whole.

There has been some initial discussion about using cell phone records to provide information that can be used to estimate urban travel behavior and at much higher frequency than the decennial Census. Already, cell phone tracking allows providers such as Google to estimate travel speeds on most roads and surface transportation systems. When combined with knowledge of system capacity, travel flows can also be estimated – on each link of a system. Tracking the actual position of cell phones at

[47] Ibid.

say 5:00 am and then 10:00 am could be used to estimate home-workplace locations and hence (to a first order approximation) an Origin-Destination matrix. This would allow firms to make know where their potential workers live, or retailers to calculate catchment areas. The use of phone records would require using raw data that already exists. Of course, this must all be done with regard to the issues of openness and privacy discussed in the previous section.

5.2.3 Tracking Urban Land Use – Concatenating Local Property Tax Parcel Data

In much of the developed world, an official and often very efficient system exists for appraising real property and collecting annual Ad Valorem taxes there upon: Property taxation. This system is always operated at the local (city, town, municipal) level. In most cases, automated databases exist for the parcels or properties within a jurisdiction that contains considerable information about the property (e.g. land area, floor-to-area ratio [FAR], age). The problem is that there are hundreds and often thousands of jurisdictions within a country collecting this data. Even within the economic definition of a "city" (e.g. Metropolitan area) often there can be hundreds of legal jurisdictions independently operating property taxation. In many instances, higher level State or National governments have imposed standardized requirements for these databases. What has yet to be accomplished is the concatenation of these data to provide a composite picture of land use at the metropolitan area level. By studying the timing and geography recent land use changes, real estate entrepreneurs could identify the most promising areas for future development.

6 Conclusions

The big data revolution is radically affecting our daily lives. We buy products recommended by online programs, watch videos suggested by streaming services, listen to music proposed by computer algorithms, and drive our cars using directions beamed to our smartphones by map applications. Crowdsourced customer ratings help us avoid bad restaurants and hotels, and faulty consumer products. Data-driven statistical models swiftly underwrite our mortgages and consumer credit. Retailers use big data to stock their shops with the products we want and to offer us the discounts we require. While most of these innovations may not be reflected in the Gross Domestic Product (GDP), many of them are making our lives better.

Governments are slowly catching up to these transformational technologies. As we have seen in this report, some of the most significant opportunities lie at the municipal level. Urban information technologies can make our cities 'smart.' Data analysis and forecast can improve outcomes in health, education, and planning. Sensors and programmed feedback loops can optimize the operations of traffic lights, street lamps, water systems, heating in and air-conditioning AC in municipal buildings, and public infrastructure. By obtaining data on user outcomes, AI algorithms can set up optimal parameters of utilization that change depending on the current environment. Opportunities are emerging to use information about citizen preferences to provide better services at a lower cost.

We are, however, at an incipient stage in the development of smart city technologies. Practitioners should thus be cautious about over-hyping and over-promotion of 'solutions' that do not solve problems in a cost-effective way. Nonetheless, the process of creative destruction over the current experimental stage will certainly leave with abundant feasible smart city technologies. We expect for innovations that substantially reduce operating and capital costs to be the first widely adopted. Technologies with clear immediate benefits in terms of quality of life will follow. Finally, technologies that improve quality of life and the productivity of cities over the long run will progressively expand as localized demonstration projects prove their benefits to outweigh costs.

We forecast that smart city technologies will reinforce the primacy of the most successful global metropolises. We reckon at least for a decade or more. These cities have the IT know-how, human capital base, and experimental attitude required. Most of the time, they also have progressive leadership. As such technologies come to full fruition, thereby improving quality of life and local productivity, these areas are better positioned to reap their full benefits. This means that real estate valuations in these areas will continue to grow robustly: especially so, in those metropolitan areas with inelastic supply due to barriers to new development. A few select metropolises in emerging countries may also leverage these technologies to leapfrog on the provision of local public services.

In the very long run, all cities throughout the urban system will end up adopting successful and cost-effective smart city initiatives, thereby diluting the first mover advantage of the leaders. Nevertheless, smaller scale interventions are likely to crop up everywhere, even in the short run. These programs are more likely to improve conditions in blighted or relatively deprived neighborhoods, which could generate gentrification and higher valuations there.

It is unclear whether urban information systems will have a centralizing or suburbanizing impact. They are likely to make denser urban centers more attractive, but they are also bound to make suburban or exurban locations more accessible.

Policies and interventions based on urban big data will develop in a period of accelerating technological change. Real estate markets will be concurrently impacted by technological innovation: new materials and construction methods, transactional platforms, smart buildings, virtual reality, and new retail models.[48] Automated vehicles will transform our transportation systems and have a deep imprint on built environment. Likewise, increasing automation and the use of IT

[48] Ibid.

systems will transform the ways in which we work and communicate. Environmental issues will come to the forefront.

After the dust settles, we should hope to see smart city technologies making our cities cleaner, leaner, and more efficient, in time to improve the quality of life of an increasingly urban global population.

Acknowledgements We would like to thank Professor William Wheaton for his feedback on an earlier draft and his valuable contribution to the section on future opportunities. We also greatly appreciate the generosity of experts for discussing the examples, research, and issues raised in this report. Finally, we would like to thank CBRE for the idea of creating this report and their financial support for research assistance.

References

Ara Andorra (2016), Primers passos d'un vehicle autonomy per descongestionar zones de vianants (http://www.ara.ad/societat/Primers-vehicle-autonom-descongestionar-vianants_0_1656434541.html)

Eells, A. and S. Fletcher (2016), Understanding Collaborative Participation: "Asthampolis" in Louisville, Kentucky (http://participedia.net/en/cases/understanding-collaborative-participation-asthmapolis-louisville-kentucky)

Florida, R. L. (2002). *The rise of the creative class: And how it's transforming work, leisure, community and everyday life*. New York, NY: Basic Books.

Froling, B. (2016), Detroit launches demolition tracker, Crain's Detroit (http://www.crainsdetroit.com/article/20160622/NEWS/160629916/detroit-launches-demolition-tracker)

Glaeser, E., Kolko, J. and A. Saiz, (2001), "Consumer City", *Journal of Economic Geography*, pp 27-50.

Goldstein, B. (2013), Open Data in Chicago: Game On, in *Beyond Transparency*, Edited by Goldstein B. and L. Dyson, Code for America Press: San Francisco.

Government of India (2015), Smart Cities: Mission Statement and Guidelines, Ministry of Urban Development (http://smartcities.gov.in/writereaddata/smartcityguidelines.pdf)

Gurstein, M. (2011), Open data: Empowered or effective data use for everyone? *First Monday*, 16 (2)

Kitchen, Rob (2014), *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage Publications: London.

Saiz, Albert and Ariana Salazar (2017). "Real Trends: The Future of Real Estate in the United States." Capital One Banking/Massachusetts Institute of Technology.

Townsend, A. (2013) *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. W.W. Norton: New York.

Willams, S., Waiganjo, P., Orwa D. and J. Klopp (2015), The Digital Matatu Project: Using Cell Phones to Create an Open Source Data for Nairobi's Semi-Formal Bus System, *Journal of Transport Geography*, 49:39 – 51.

Williams, S. (2013), Beijing Air Tracks: Tracking Data for Good. in *Accounting Technologies: Tools for Asking Hard Questions,* Edited by Offenhuber, D., Schechtner, K., AMBRA, Austria: Vienna. (http://civicdatadesignlab.org)

Yongling, L., Yanliu L. and S. Geertman (2015), The development of smart cities in China. Working Paper, Computers in Urban Planning and Urban Management (http://web.mit.edu/cron/project/CUPUM2015/proceedings/Content/pss/291_li_h.pdf)

Social Media-Based Intelligence for Disaster Response and Management in Smart Cities

Shaheen Khatoon, Amna Asif, Md Maruf Hasan, and Majed Alshamari

1 Introduction

With the rapid urban development and scarcity of natural resources, cities worldwide are facing challenges in maintaining sustainable development and improving the quality of life of their citizens (Angelidou 2015). The concept of smart cities has emerged to address those challenges by focusing on the integration of human, collective and technological capital in urban development to achieve a better quality of life and sustainable economic, social and environmental development (Chourabi et al. 2012). To achieve that goal, smart cities leverage the capabilities of technological infrastructures, populations, and institutions and revolve around four main facets: (1) Sustainability; (2) Quality of life; (3) Urban development; and (4) Intelligence (Dhingra and Chattopadhyay 2016).

To tackle the intelligence aspect of smart cities, one of the key capabilities is responding to man-made and natural disasters in a timely and effective manner to protect the life and property of citizens. The capability required continuous monitoring of city-related infrastructure and data as intelligence for streamlining the city's emergency operation. For instance, data and intelligence about a crisis can be used by emergency response organizations of a city for coordinating crisis response activities and decision making. Depending upon the scale of a disaster, response organizations may include government, public authorities, commercial entities, volunteer organizations, media organizations, and the public. In a crisis, these entities work together as a single virtual entity to save lives, preserve infrastructure and community resources, and reestablish normalcy within the community. For the

S. Khatoon (✉) · A. Asif · M. M. Hasan · M. Alshamari
Department of Information System, College of Computer Science and Information Technology, King Faisal University, Al Hofuf, Saudi Arabia
e-mail: ssyed@kfu.edu.sa

P. M. Pardalos et al. (eds.), *Artificial Intelligence, Machine Learning, and Optimization Tools for Smart Cities*, Springer Optimization and Its Applications 186,
https://doi.org/10.1007/978-3-030-84459-2_11

emergency operation, the response network needs to gather situational information (e.g., state of the civil, transportation, and information infrastructures), together with information about available resources (e.g., medical facilities, rescue, and law enforcement units). One of the challenges of gathering situational information is the collection of real-time data related to human life and infrastructure damage.

In the past, response organizations were using the traditional emergency management systems in combination with remote sensing data interpretation techniques in order to manage emergencies. An example of such a system is the Copernicus Emergency Management Service (CEMS) (Copernicus 2018), managed by the European Commission (EC) and established in 2012 for forecasting floods and fires with the use of optical satellite imaginary. TerraSAR-X (Mason et al. 2010) is another example using synthetic aperture radar satellite-based images to assess the damage after a flood. However, recent studies showed significant operational delays in services using such systems due to delays in collecting and analyzing satellite images. For example, CEMS takes 48–72 h to analyze the image after a disaster has occurred (Schnebele and Cervone 2013). Besides speed, another limitation of such systems is the information accuracy. Studies showed that the accuracy of an optical satellite-based damage assessment after an earthquake is approximately 65% (Dell'Acqua and Gamba 2012), whereas the accuracy of a synthetic aperture radar satellite-based damage assessment after a flood is around 75% (Mason et al. 2010). The speed and accuracy at which information about the crisis flows through the disaster response network has the potential to revolutionize the existing crisis management systems.

Recently, the ubiquitous connectivity and proliferation of social networks opened up new crisis management opportunities through crowdsourcing. Researchers have started developing intelligent platforms powered by crowdsourcing and emerging technological innovations, such as cloud computing and big data analytics (BDA) to build systems for getting insight from user-generated content on different social media platforms (Dell'Acqua and Gamba 2012; Huyck and Adams 2002; Schnebele and Cervone 2013; Starbird and Stamberger 2010). Among these applications, mobile crowdsensing for the collection of crisis-related information has attracted much attention in academic and industrial forums. In this approach, the general public collaborates with the Emergency Operation Center (EOC) of a city in gathering situational information related to phenomena of interest (e.g., incident location, disaster-related losses, and needs of victims) and helps in organizing disaster response-related activities. In modern days, whenever any crisis occurs, the national security agencies, civil defense, emergency and humanitarian professionals, media outlets, and the general public join hands and collaborate using modern technologies such as social media.

The potential use of social media caught the crisis management research community's attention when close attention was given to social media by the crisis management organization Red Cross in its Emergency Social Data Summit held in August 2010 (Cross 2010). Since then, the research community has started exploring how to monitor, analyze, and display data extracted from social media. A recent survey (Yu et al. 2018) on the role of big data in disaster management shows that social media has been used as the most prominent information source in various disaster management phases, such as long-term risk assessment and reduction, forecasting and prediction, monitoring and detection, early warning, and post-disaster coordination and response. However, extracting meaningful information from such a large, diverse, dynamic, but potentially useful data source is a big challenge that is just beginning to be addressed by the research community.

One such crowdsourcing tool is Ushahdi (Okolloh 2009). It was initially developed to visualize crowdsource reports of violence in Kenya after the post-election violence in 2008. Since then, it is expanding and has become an excellent example of crowdsourcing for raising awareness on different social issues. This platform collects users' generated data through SMS, email, and Twitter, and then it visualizes it using maps and charts.

"MicroMappers" (Meier 2015) is a relevant example of a crowdsourcing platform, which collects images and texts from social media and classifies them into predefined disaster categories for the effective response of humanitarian organizations. SensePlace2 (Mac Eachren et al. 2011) was developed to collect crisis-related tweets with place and time information and then visualize this information using maps and time plots.

All of the systems mentioned above are helpful to identify the different disaster types and their locations. However, these systems are limited to a single functionality of visualization with a specific information need. Moreover, they only use a single source of data such as text or images and collect data from one or two social media sites, respectively. Furthermore, a crisis response process requires an end-to-end Information Technology (IT) solution for organizing crisis response activities include crisis reporting, damage and need assessment, measures undertaken to protect life and property immediately before, during, and immediately after disaster impact.

The shortcomings of the existing sensors' based disaster management system and the potential of improvements in social media-based platforms led us to propose dynamic social media analytics and crowdsourcing framework for assisting/improving a city's EOC. In this chapter, we investigate a cloud computing-based big data framework that will enable us to utilize heterogeneous data sources and sophisticated machine learning techniques to gather and process information intelligently and provide emergency workers useful insights for making informed decisions. Such a comprehensive framework will help a city develop comprehensive Disaster Risk Management capability to automatically predict hazards, early warning, risk assessment, and risk mitigation, including coordination of emergency activities and evacuation.

To the best of our knowledge, the proposed system will be the first to develop a next-generation IT solution for social media analytics focusing on disaster informatics. The system will be scalable to real-time situational awareness of any large-scale event, not just restricted to man-made and natural disasters. Additionally, an advantage of the proposed approach over the already existing system is the use of multilingual-multimodality data over several social media channels, which will increase the quality and quantity of data leading to an improved results analysis.

2 Social Media Analytics System for Emergency Event Detection and Crisis Management

This chapter aims to investigate a cloud-based integrated solution for disaster and emergency management using social media analytics. The main thrust is augmenting existing sensor-based Disaster Risk Management (DRM) systems with social media capabilities by keeping the public in the loop (human sensors). The solution will allow relevant disaster management authorities to integrate and access data from several internet-based social data sources and apply semantic analysis to generate actions to be executed in response to presented contents. The result will be used by relevant emergency monitoring and disaster management agencies for emergency response, early warning, risk assessment, and risk mitigation, including coordination of emergency activities.

Figure 1 depicts the overall architecture, which comprises components for data acquisition, storage, management and analysis, and graphical user interfaces for the emergency response operators and the crowd. We have identified the following key components in building next-generation IT solutions for disaster response and management centered around disaster events.

Event Extraction and Interpretation: Automated event detection capabilities are required to extract events from streams of all modalities, i.e., text, audio, and video. Multimodal extraction, fusion and assimilation technology that enables extraction and interpretation of disaster related information from multiple social media channels.

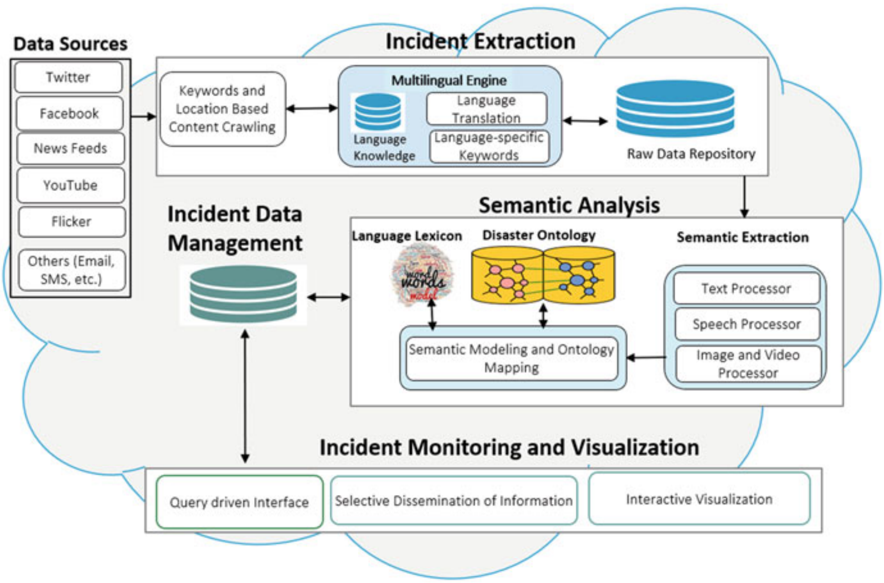


Fig. 1 The abstract architecture of social media-based incident detection and monitoring system

Semantic Analysis: The data shall be transformed in a format suitable to apply subsequent semantic analysis and stored in the event database following preprocessing. Robust technologies are required to extract semantic information from raw data streams and go all the way to high-quality extracted information for decision-makers in the forms most appropriate for their various tasks. Automated capabilities are required to process the various pieces of information, such as images, audio and video feeds, text feeds, etc., and also integrated and fuse information coming from multiple disparate sources. We will develop a disaster ontology to map metadata derived from social data to the ontology matching concept. Ontology creation and alignment are the immediate future work under this research project. At this stage, we assume that the ontology will have classes for different disaster types and information about relevant relief organization and the location of those organizations.

Event Data Management: Technologies that support modeling, storing, querying and indexing, situational information. We need data management capabilities for storage and structured querying of the information collected, processed, and integrated into the ingest phase. The information obtained about the unfolding situation and status should be queryable in a structured manner.

Monitoring and Visualization: Capabilities to effectively monitor and visualize crisis-related social media contents for selective dissemination of information with disaster response network; visualize incident-related information such as incidents graphs showing a relationship between events and statistics about the different incidents; and query and navigation interfaces to allow users to filter through the detail of incidents and events such as location, damage etc.

We now describe techniques related to each tier in the proposed architecture to provide an overview of the methods integrated into each component.

2.1 Incident Extraction and Representation

In the context of emergencies, incident extraction refers to the task of discovering a new event by continuously monitoring raw data streams. Automated event detection capabilities are required to extract events from streams of all modalities, i.e., text, audio, and video. Most large social media platforms provide programmatic access to their content through an Application Programming Interface (API). APIs allow data collectors to express an information need, including one or several of the following constraints: (1) a time period; (2) a geographical region for messages that have GPS coordinates or (3) a set of keywords that must be present in the messages.

The workflow can either be triggered automatically upon detecting an event matching with information need or an operator manually on deployment. To start the crawler, several parameters need to be specified.

The location-based crawler required a predefined area of interest and time window size for all social media networks configured for crawling. The location will be provided using static location coordinates (i.e., longitude, latitude) through Google API.

The keyword-based crawler starts with the specific search terms or uses predefined terms stored in the language database to initiate the search in the keyword search. The crawler will then start searching the contents matching with search terms. Keyword-based filtering can be used to continuously monitor the activity in social media networks concerning natural disasters. The frequency of particular words associated with disasters can be temporarily analyzed to detect a new or upcoming disaster.

The multi-language component's goal is to provide the capability of crawling the content uploaded in different languages. The system will provide a language translation service using Google/Microsoft language translation APIs to translate posts in the target language and store language-specific keywords in the language knowledge-base. Upon setting up these parameters, the application starts receiving data from one or more social media sites, such as Twitter, Facebook, YouTube, and news feed.

The content from one or more sources may contain various forms, including text (posts, comments on blog posts, news, etc.) and media uploaded with the post (image and video contents with rich metadata information such as location and text) and videos. Following the data collection and translation, the data crawled is transformed into a suitable format to apply preprocessing techniques on text, image, and video, before applying semantic analysis. Data preprocessing techniques are described under the relevant section of each modality.

2.2 Text Processor

This research area focuses on developing robust text analytic approaches for the extraction of information from text, starting with raw reports/posts or documents and providing extracted information of high quality and reliability to the end-users. One key aspect of effective disaster response is up-to-date information on the area of interest (Seppanen and Virrantaus 2015). To this end, developing useful information processing tools to identify the disaster type, timing, and locations is critical (Seppanen and Virrantaus 2015).

The use of social media data for disaster management has led to creating a variety of text analytics techniques for information processing (Zhang et al. 2019). There are three main streams of research in the existing literature to leverage social media posts for enhanced situation awareness during disaster response and recovery: (1) text classification, (2) geographical estimation, and (3) information extraction.

2.2.1 Text Classification

The first stream of research focuses on supervised and unsupervised algorithms to categorize extracts in one or more disaster-related classes. Under supervised algorithms, multiple machine learning classifiers have been used to classify social media data with disaster-related topics, such as Naïve Bayesian classifier (Imran et al. 2013), Convolutional Neural Networks (Huang et al. 2019), and Generative Adversarial Network (Dao et al. 2018). A supervised algorithm can be used when a set of example items in each category is given. However, provided the dynamic nature of social media posts, it is impossible to predetermine labeled training set for each type of disaster.

Since supervised approaches for text classification require human-annotated labels, the use of the unsupervised approaches has started increasing in the last decade or so. Unsupervised methods seek to identify and explain important hidden patterns in unlabeled data. Well-known unsupervised learning methods used in text analytics are topic modeling and clustering. Topic models reveal the latent topics by identifying words' occurrence patterns in a document. There are many studies that use topic modeling to extract topics corresponding to disaster events. For example, Ragini et al. (2018) used a set of keywords that refer to danger, such as "trapped, stranded, help, save, rescue, struck, caught" to extract relevant information from disaster-related tweets. Mishler et al. (2015) used the Structural Topic Model (STM), a variant of LDA, to capture temporal changes in the Ukrainian crisis. Kireyev et al. (2009) used basic LDA on Twitter data from two earthquakes in America and Indonesia to uncover the most prominent topics during these two natural disasters. They emphasized dynamic corpus refinement to overcome the content sparsity problems in a short text by training the topic model on large external corpora to enrich the short representation of text.

These techniques are proven useful in detecting disaster events using posts shared on social media. However, in the context of effective emergency response, the extracted information should fulfill the needs of a disaster response network, targeting specific information needs. Many response organizations have predetermined information needs. For example, in a building collapse, a field worker might require detailed information about the event, while a rescue team may only need to know the location and number of injured people in the vicinity of the catastrophe. By considering the information need of a specific organization, in this project, we focus on employing state-of-the-art AI techniques to analyze and understand useful information for humanitarian decision-makers and responders.

2.2.2 Location Estimation

Another stream of studies has focused on assessing disaster damages over different locations for disaster-related posts. By doing so, the unfolding of events can be mapped to a particular location. The majority of the existing social media platforms retrieved disaster-related tweets across different geographical locations to identify spatial patterns of a crisis region. Most of the current systems primarily employed geotagging to automatically extract geocoordinate (latitude/longitude) from posted content metadata. For example, SensePlace2 (MacEachren et al. 2011) extracted geotags using explicit geolocation metadata in the tweet, location of users' posting, and place-based hashtags.

Similarly, Twitcident (Abel et al. 2012) used a geography boundary box (longitude/latitude of a specific location) to gather geotagged tweets only and employs classification algorithms to extract messages about small-scale crisis response, such as music festival and factory fires. Tweak the Tweet (Starbird and Stamberger 2010) used tweet syntax to encourage content originators to markup location-based keywords, which assist in filtering and classification of emergency-related information. Using the relevant hashtags, people can craft more relevant geotagged content to capture the situation in disaster-affected areas. However, the author concluded that the content originator had not widely adopted syntax. TweetTracker (Kumar et al. 2011) parses a Twitter feed to extract location-specific keywords and place-based hashtags to monitor and analyze crisis regions to improve situational awareness. Although many disaster management systems support geotagged capabilities, the availability of explicit location metadata from social media posts is limited. Situational information published by reporters and individuals tends not to attach the geotag in their posts due to privacy concerns. In particular, only 1% of the posts include machine readable location metadata (Malik et al. 2015).

To further enhance location insights from social media posts, many studies used the concept of volunteered geographic information (VGI), which focuses on utilizing the potential of crowdsourcing to generate useful geographic information to track and group geotagged social media posts on the location map. For example, MicroMappers (Meier 2015) employed volunteers to manually geotag messages into different humanitarian categories across different locations. The tool then constructs a location map of these categories that can be sent to response agencies providing the up-to-date status of affected areas. Ushahidi (Okolloh 2009) used crowed to manually tag posts from a wide range of resources and geo-visualize crowdsource reports to increase situational awareness of the affected region. CrisisTracker (Rogstadius et al. 2013) integrated crowdsourcing techniques to annotate rapid streams of unstructured tweets with metadata and cluster related stories based on their lexical similarity. It then utilized volunteers to verify and further curate stories by adding additional metadata, removing duplicate stories and/or merging similar stories. Zook et al. (2010) provided a recent overview of VGI/social media applied to crisis management with a focus on Haiti.

Despite the popularity of crowdsourcing techniques to curate location information for further interpretation, the effectiveness of such systems is highly dependent on the motivation and size of the crowd. Furthermore, none of the existing crowdsourcing frameworks utilized the inference techniques to automatically discover location from huge data and disseminate results to relevant authorities. Understanding the social media attention to a crisis-affected region requires detecting geographical information of the location in the posts. Many social media platforms, such as Twitter, allow users to share geographic information in the form of physical

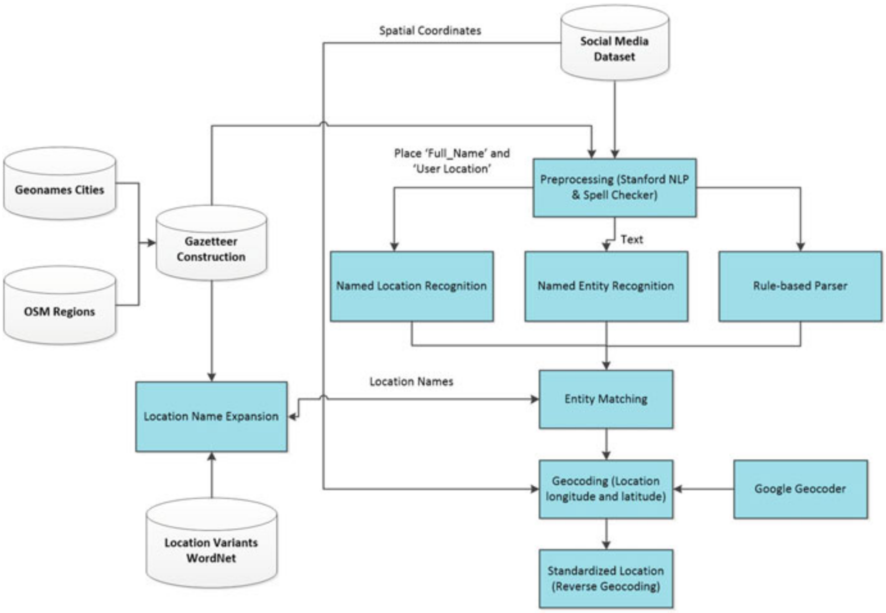


Fig. 2 Location extraction from Twitter posts

addresses specified in user profiles; however, users' tend not to share their locations due to privacy concerns. Therefore, a small fraction (less than 1%) of social media posts usually have location information (Malik et al. 2015). However, some studies (Fan et al. 2020; Liu et al. 2013; MacEachren et al. 2011; Middleton et al. 2013) have discovered that locations mentioned in the posted content can provide an additional opportunity to complement the limited geotagged posts. By considering the above problems, in this work, we selected Twitter as a case study to infer tweet geolocation from geo-coordinate, user location tweet place-field, and location mention in tweet content as shown in Fig. 2.

Twitter provides an option to enable the exact geolocation (geocode) of users on mobile devices that contain GPS coordinates, i.e., latitude and longitude. These coordinates can be directly mapped into a valid spatial representation using geocode services. There are several commercial geocoding services (Google Geocoding API,[1] OpenStreetMap Nominatim,[2] and Bing Maps API[3]) based upon an underlying map database, which can take well-formatted location descriptions and return map references to them.

[1] https://developers.google.com/maps/documentation/geocoding
[2] http://wiki.openstreetmap.org/wiki/Nominatim
[3] https://www.microsoft.com/maps/choose-your-bing-maps-API.aspx

Tweet content represents the actual tweet in text format. To extract location mentions in the tweet text, we tried several publicly available Named-Entity Recognition systems, including NLTK2, a recent version of Stanford NER (Finkel et al. 2005), and a system by (Ritter et al. 2011) to extract location mention in the tweets. However, user location, place full name, and location mention in tweets are unstructured content, which needs location parsing and location disambiguation prior to geocoding. We used a geolocation algorithm (GeoLocator v1.0)[4], which contains both geoparser that extract locations and a geo-coder that assigns latitude and longitude to each location.

However, due to either unsatisfactory results (e.g., most of them need proper capitalization to detect location names) or not enough computational speed to deal with a large dataset, we could not directly use these tools. Instead, we have adopted a gazetteer-based approach and used Nominatim, a search engine for OpenStreetMap (OSM) data. Specifically, we used the Nominatim tool to perform two operations: (1) geocoding—the process of obtaining geo-coordinates from a location name, and (2) reverse geocoding—getting a location name from geo-coordinates. Furthermore, we used WordNet to identify location variants to map each location into its standard format.

By using the Google Maps Geocoding API, the coordinates of the recognized location entities can be saved. Accordingly, we can use the latitudes and longitudes for the identified locations to locate a post on a geographic map to estimate the density of posts across various locations for different categories of events.

2.2.3 Information Extraction

The task of Information Extraction (IE) involves automatically extracting structured information from unstructured (e.g., plain text documents) or semi-structured (e.g., web pages). The most common IE task is Named Entity Recognition (NER), which consists of detecting regions of a text referring to people, organizations, or locations (Liu et al. 2013; Ritter et al. 2011).

In the context of disaster-related posts, IE can be used to create incident reports by extracting facts from social media posts written in natural language. For example, "10 injured and 4 dead in a traffic accident in Riyadh" can be normalized to transform into machine-readable record such as {< affected-people = 10, report-type = injury, location = Riyadh, Saudi Arabia>}, {<affected-people = 4, report-type = casualty, location = Riyadh, Saudi Arabia>}. These normalized records can be easily filtered, sorted, or aggregated.

Such systems can be built by specifying a 'schema' for the facts to be extracted along with semantic information and properties about the different attributes of events. The user may also provide some logical deductive rules that state what event slot should be filled with entities. Many 'off-the-shelf' language processing

[4] https://github.com/geoparser/geolocator

tools are available to develop powerful natural language processing applications, for example, text analytics tools such as GATE (Cunningham 2002), semantic parsers such as Shalmaneser (Erk and Pado 2006) and the Stanford NER Parser (Klein and Manning 2002).

We are developing a text processing platform with automated capabilities that extricate information by combining the above task under a single event-centered platform. The system involves several phases, such as first classify which post could contain the disaster-related event in the first place utilizing supervised and unsupervised methods, then grouping relevant posts using clustering approaches, and finally extracting entities by filling slot values utilizing named entity linking semantic technology.

2.3 Image Processor

The images are an integral part of social media posts and sources of exhaustive information about the events. The usefulness of visual contents of social media posts in crises can be utilized in many ways, such as to collect contextual information (e.g., the current status of transportation or infrastructure), information about available resources (e.g., food supply, medical aid, water shortage), information regarding damage (e.g., damage severity levels) and the awareness status (e.g., the warning). Such information can be collected by looking at the visual contents to identify the type of information an image carries. However, it is challenging to extract comprehensive information from the massive number of social media data, as it is not possible to perform this task manually. Therefore, it is required to develop an approach that allows automating the disaster-related information filtering process.

We developed machine learning and related image recognition algorithms for image-based event detection for enhanced situation awareness. There are two primary tasks for extracting semantic information from images, as shown in Fig. 3. (1) image acquisition and labeling and (2) image processing pipeline. In the following section, we will describe methods integrated into each component.

2.3.1 Image Acquisition and Labeling

Various types of images related to different hazard types such as collapsed buildings and infrastructure, flood, hurricane, etc., can be collected through a web search, social media platforms (Flickr, Twitter, Instagram, etc.,), and publicly available image catalogs. Towards developing a successful image processing model, it is essential to annotate the image dataset correctly with precise labels to identify different types of information related to a crisis. Although few studies exist on information extraction from disaster-related images in the literature, however, they focus on a specific information type such as 'severe damage', 'damage', or 'no damage'

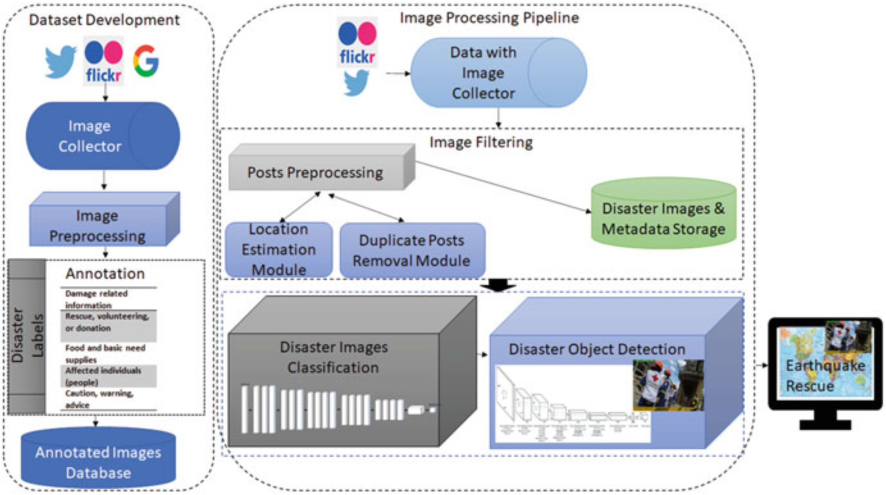


Fig. 3 Image processing model for disaster type classification and disaster related objects detection for emergency response

(Alam et al. 2018a; Barz et al. 2019; Murthy et al. 2016; Olteanu et al. 2015), first-aid activities such as rescue, volunteering, or donation (Alam et al. 2018a; Olteanu et al. 2015), the need of supplies such as food and basic needs (Murthy et al. 2016; Olteanu et al. 2015), people who are affected from the crisis (people) (Alam et al. 2018a; Murthy et al. 2016; Olteanu et al. 2015), warning information (Olteanu et al. 2015), and information regarding social activities during the crisis (Gaur et al. 2019). There is a tremendous potential for further improvements in the delivery of crisis-related information to relevant response authorities through the development of a single platform addressing different information needs. Therefore, we have developed a comprehensive disaster taxonomy targeting the information need of a specific responder. The taxonomy consists of six different disaster categories, including: (1) damage related information, (2) rescue, volunteering and donation, (3) food and basic needs supplies, (4) affected individuals, (5) caution, and (6) effect on social activities.

To map the disaster related images on the proposed categories in the taxonomy, image annotation is needed. There are some existing open-source datasets consist of disaster-related annotated such as CrisisMMD (Alam et al. 2018a), European Flood 2013 Dataset (Castillo 2016), and Fire Dataset (Saied 2020). These datasets are useful in classifying the images into various disaster types such as earthquake, flood, hurricane, and fire etc. However, the organizations working for the disaster management activities require specific information from an image for the appropriate response. To further facilitate disaster operations, we can annotate disaster-related objects from a given image using a 2D bounding box (Dwibedi et al. 2016) technique. Several platforms are available for image annotation, such as CrowdFlower,[5] Amazon Mechanical Turk[6], and AWS rekognition.[7] These platforms allow users to annotate a specific object by drawing a bounded box around the objects within the images using the human workforce. To develop the dataset, we first downloaded images from publicly available resources (Flicker, Google, Twitter) using the specific disaster-related keywords. We then performed basic preprocessing such as resizing and removing the duplicates. In the next step, we annotated images by drawing bounded box around disaster-related objects using opensource tool LabelImg[8] and labeled them according to proposed labelling scheme/taxonomy.

2.3.2 Image Processing Pipeline

We developed an image processing pipeline comprised of images collector, image pre-processor, image classifier, object detector, and visualization modules for a detailed analysis of imagery content.

At first, images were downloaded using Twitter and Flicker) and various preprocessing techniques are applied to the raw images. In the case of Twitter, only the tweets contained media URLs are selected, and images are extracted using *Madia* class from tweeter API with attributes of *id* and *media_url*. The data collection from the Flickr platform is done by using Flickr *search* API. In the first preprocessing step duplicates are detected and erased from the data set. The remaining images are used for further interpretation using machine learning and related image recognition algorithms. Besides the photos, the spatial location of images is an essential component for mapping the results. The location extraction task is challenging due to limited data with explicit location information; therefore, the messages/social media post metadata is used for the geo-localization using location estimation techniques discussed under the text processor module. In the previous studies, location is estimated using selection of tweets with media and fine-grained localization (Francalanci et al. 2017). The geo-localized images are extracted from social media data to support emergency response and used in the image processing pipeline to analyze area-specific images and their usefulness (Ionescu et al. 2014; Murthy et al. 2016; Peters and de Albuquerque 2015). Keyword filtering is another way of geo-localization where the Euclidean distance is calculated between the users and the hazard location. (Peters and de Albuquerque 2015).

After location extraction and duplicate removal, image classification techniques are applied to classify the images on predefined categories. A range of classification techniques from binary classifier such as disaster and non-disaster images to multi-classifier on disaster types such as earthquake, flood, hurricane, etc., humanitarian

[5] http://faircrowd.work/platform/crowdflower/

[6] https://www.mturk.com/

[7] https://aws.amazon.com/rekognition/?nc2=h_ql_prod_ml_rek&blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc

[8] https://github.com/tzutalin/labelImg

categories (Nguyen et al. 2017b), situational awareness (Peters and de Albuquerque 2015), damage severity (Alam et al. 2018a; Nguyen et al. 2017a), flooded and dry images, and water pollution (Barz et al. 2019) discussed in the previous literature. For the classification of images on certain damage types, various deep-learning neural networks, specifically convolutional neural networks, can be applied to gain accurate results.

For the image classification task, we used the Google Tensorflow library for building a machine learning pipeline. It allows transfer-learning, meaning that it is not required to develop a model from scratch but to build on a previously trained model and to develop only a new final layer. Various transfer learning architecture such as AlexNet (Iandola et al. 2016), GoogleNet (Szegedy et al. 2015), VGGNet (Simonyan and Zisserman 2014), ResNet (with 50, 101, and 152 layers) (Targ et al. 2016) are explored to select the best classification model. Among these models, fine-tuned VGG-16 network achieved the best performance with 95% accuracy. However, the performance of different disaster classification deep neural networks depends on the structure of a neural network, dataset image quality, the size of the dataset, the type of disaster images, and the quality of image annotation. The results of this study are similar to the previous research where VGGNet (Alam et al. 2018b) achieved the best performance, however, the classification results are only available on limited scenarios of disaster and datasets. Therefore, the challenge is to investigate the deep learning neural network to attain the best performance in classifying as many types of disasters as possible.

The disaster-related social media images can be further explored to discover additional details of the damage within an image, such as resources, rescue needs and services, and warning signs, which require object detection techniques. The previous studies are very limited on discovering disaster-related objects from the images. One of the main focuses of our research is the investigation of disaster-related objects and detection from the images. This will allow utilization of the social media posted image data in effective disaster emergency response. In the proposed pipeline, after classifying an image into the relevant category, object detection techniques are applied. We used YOLOv4 (Bochkovskiy et al. 2020) object detection model to recognize the disaster-related objects. The identified objects are then fused with the disaster class identified in the previous step to narrate the meaning from the image and sent to appropriate emergency response workers. This will help filter and transfer a huge number of social media images efficiently to the specific authority responsible for disaster management activities. In the previous research, object detection is applied to find common objects in the images in emergencies. For this purpose, the techniques of the Faster R-CNN (Ren et al. 2016) trained on COCO dataset (Lin et al. 2014) are applied (Chino et al. 2015). However, there are limited contributions to the area of disaster-related object detection from social media images. Many object detection algorithms based on region proposal-based methods such as R-CNN (Girshick 2015), SPP-net (Purkait et al. 2017), Fast R-CNN (Girshick 2015), Faster R-CNN (Ren et al. 2016), R-FCN, FPN, and Mask R-CNN (He et al. 2017) and regression/classification based methods such as MultiBox (Erhan et al. 2014), AttentionNet (Yoo et al. 2015), G-CNN

(Najibi et al. 2016), YOLO (Redmon et al. 2016), SSD, YOLOv2, YOLO3 (Farhadi and Redmon 2018), YOLO4 (Bochkovskiy et al. 2020), DSSD, and DSOD (Zhao et al. 2019) discussed in the literature for object recognition. However, we investigated one of the best performing object detection algorithm YOLO. Further research is needed to investigate the efficient object detection methods for identifying the disaster-related objects to understand the context and the details about the situation.

2.4 Audio-Video Processor

Information shared on social media is often not unique but available in multiple versions. As part of the proposed model, we are developing approaches for extracting information from a single modality and extracting and fusing information from multimodal data. The research focuses on developing approaches for the interpretation and analysis of audio and video data.

Video-based event detection can be done using computer vision technology. In the context of emergency event detection, we tend to identify objects and action concepts. Examples of object concepts for crises can be fire, car accidents, building collapse, etc. Object concepts can be detected by evaluating different object detectors on each video frame and aggregating the results temporally over the video. State-of-the-art object detectors based on deep neural network architects can provide thousands of object detectors, pre-trained over large image datasets such as ImageNet (Krizhevsky et al. 2012). Examples of action concepts include fighting, running, crowd pushing, etc. Detection of action concepts can be achieved by extracting low-level dynamic features, including dense trajectories (Wang and Schmid 2013) and STIP (Laptev et al. 2008), as well as static features (Dalal and Triggs 2005). In the proposed architecture, objects and concepts from the selected videos will be extracted into a significant number of still photos. In the next step, image analysis techniques will be applied to classify images extracted from the video, as explained in the previous section.

Audio-based event detection requires automated capabilities for feature extraction from low-level signals and robust speech recognition technologies. In the context of emergency event detection, audio-based concepts include sounds of gunshots, explosions, etc. Audio concepts can be detected by learning classifiers on MFCC features (Mel-Frequency Ceptral Coefficients) from the audio track in a way similar to (Logan 2000). Automatic speech recognition provides a stream of words (text) corresponding to a speech in an audio or video. We are developing an end-to-end speech recognition engine that displays words as if they are spoken, and respective text analytics techniques (explained in the previous section) will be applied to identify disaster-related cues.

2.5 Ontology for Disaster Management

Disaster response and planning usually involve various organizations addressing different aspects of a problem in a coordinated fashion. Law enforcement authorities and humanitarian agencies often coordinate their operations using manual interactions to save lives and minimize property and infrastructure damage. Recently, social media are flooded with a huge amount of user-generated information related to natural disasters or national emergencies at a scale that is beyond the capacity of manual human processing. Automation is needed in both identifying the misinformation as well as integrating the fragments of correct information from various sources so that the relevant stakeholders can act on such information accordingly.

One of the major challenges in crisis management is about developing solutions that can seamlessly bring together relevant information from all types of sources (text, images, audio-video, and geo-spatial) to support crisis-related activities amongst the relevant stakeholders (from the field to the situation room, and amongst situation rooms as well as workers in the field) during crises. In addressing such challenges, an ontology-based knowledge representation to validate and integrate social media contents is inevitable in facilitating the seamless flow of information with relevant agencies and stakeholders working together in emergency and crisis management.

In an ontology-based approach to disaster management, first, we need a unified representation of standard vocabularies that describes the concepts as well as the attributes, constraints, and relationships among the disaster-related concepts. Such an explicit specification of conceptualization of the disaster domain is known as the Disaster Ontology. We also utilized various natural language processing tools to semi-automatically maintain the ontology using relevant knowledge extracted from plain texts (e.g., social media postings and news reports) (Khatoon et al. 2020). Tools are also being developed to exchange or integrate information between our ontology and heterogeneous databases located at the sites of the law-enforcement, emergency response, and humanitarian agencies.

We integrated our disaster ontology with a cloud-enabled crowdsourcing platform, Ushahidi.[9] The integrated system helps users creating a crisis reporting interface augmented with the domain knowledge. The disaster ontology also helps us with identifying the relevant authority or agency and a series of relevant actions (workflow) that are crucial in coordinating the emergency response operation effectively. The disaster ontology is equally useful in data integration with heterogeneous databases maintained by diverse stakeholders coordinating the emergency response operations in a seamless manner.

[9] https://www.ushahidi.com

2.5.1 The Ushahidi Crowdsourcing Platform

Ushahidi is an open-source crowdsourcing platform which is in continuous development since 2008. Ushahidi has been deployed by various communities and international aid organizations to support disaster handling and management through crowdsourcing. The current cloud-based implementation of Ushahidi is extremely flexible as it can be easily integrated with popular social media platforms, external databases as well as SMS and email gateways. However, such flexibility comes with a cost. Most of the configuration, integration, and day-to-day operation on Ushahidi are done manually. For example, to gather information about a particular type of disaster, the Ushahidi Administrators need to setup reporting forms (templates) along with a series of tasks (workflow) manually. The comprehensive reporting of different types of disasters requires specialized human expertise with relevant domain knowledge, which is not always readily available. Incomplete or inaccurate reporting often leads to confusion and delays in emergency response. For example, unlike COVID-19, in reporting a viral pandemic like MERS (Middle Eastern Respiratory Syndrome) outbreak, it is crucial to include information such as proximity with Camel or Camel-milk consumption related information. Subsequently, as part of the emergency response workflow, together with the notification of the public health authorities, the local veterinary department must also be alerted for the appropriate handling of the infected animals. Similarly, for a traffic accident involving humans' bodily injury at a particular geographic area, the relevant paramedics will be informed together with the traffic police and insurance agents as appropriate.

2.5.2 Integration of Disaster Ontology with Ushahidi Platform

The cloud-based implementation of the current version of the Ushahidi crowdsourcing platform is modularized into two distinct modules – (1) Platform API: Crowdsourcing backend implemented as Kubernetes with native APIs and Webhooks, and (2) Platform Client: Crowdsourcing Frontend for data gathering, analysis, and visualization. With Ushahidi, we developed and integrated an owl-based disaster ontology using Protégé ontology editor[10] and the opensource Pallet reasoner plugin (capable of performing description logic-based reasoning and inference). The disaster ontology along with the natural language processing and annotation tools we developed, are integrated with the Ushahidi codebase. The current implementation and deployment of the disaster ontology enhanced Ushahidi crowdsourcing platform focuses on demonstrating the following features as a proof-of-concept basis:

1. Generate Reporting Forms (or Templates) semi-automatically for any disaster and emergency using the domain knowledge from the disaster ontology.

[10] https://protege.stanford.edu/

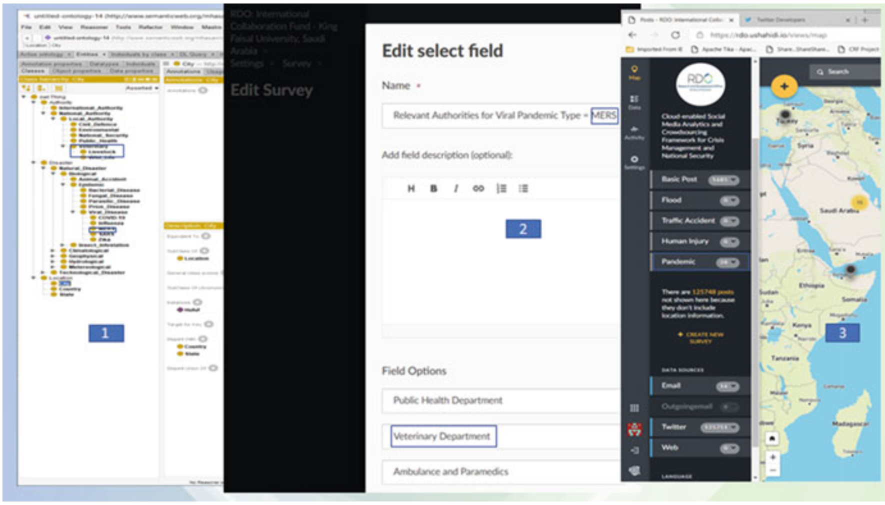


Fig. 4 Ushahidi crowdsourcing platform with disaster ontology, (1) Disaster ontology showing the key classes and attributes, (2) Editable Ushahidi report form generated with the help of the ontology, and (3) Ushahidi crowdsourcing platform showing the reported disasters on the World Map

2. Generate Emergency Response Workflows automatically with necessary information inferred from the disaster ontology.
3. Integrate information in our disaster ontology with the heterogeneous data sources located at the site of the humanitarian organizations and emergency response authorities using webhook and other API seamlessly.

As shown in Fig. 4, with the help of the disaster ontology, any Ushahidi Administrator is guided to generate an appropriate *Reporting Form* to gather data from users via the Ushahidi web interface directly (or through an appropriate *Template* via the Twitter, email, or SMS gateway where semi-structure data is already annotated using natural language processing tools automatically). For a viral pandemic like MERS, camel-related attributes are included automatically as relevant. In case of other viral pandemic, such attributes are not relevant and will be automatically excluded. The figure also shows the auto-generated *Emergency Response Workflow* that includes information derived from the disaster ontology through inference (e.g., the contact details of the Veterinary Department of the relevant geographic area is extracted from the disaster ontology using the location attribute and the type of viral pandemic).

We are continuously enhancing the disaster ontology to integrate knowledge related to various disasters and the related workflow. We are also developing and enhancing necessary tools using natural language processing techniques and machine learning algorithms to maintain the ontology semi-automatically so that we can facilitate automatic annotation and detection of named-entity and disaster-specific events, topics, and workflow, etc. Integration of the information in our disaster ontology with heterogeneous data sources (e.g., HDX, the Humanitarian Data Exchange server[11]) are also currently in progress.

2.6 Incident Monitoring and Visualization

Incident monitoring and visualization are concerned with developing an intuitive visual interface for end-users. End-users such as emergency operations managers would like to see visual patterns and trends for incident type over time and space. In this regard, we are developing a comprehensive dashboard to facilitate information visualization of multi-model data. The dashboard will provide integrated and fused information from multiple sources and provides querying, filtering, and analysis functionality.

Previous studies have used different ways to analyze crisis-related information by targeting specific information needs visually. For example, Chae et al. (2014) focused on how the public follows evacuation orders in an emergency situation. They investigated people's movement patterns by comparing spatiotemporal patterns using visualization that helped analysts understand how users react during an emergency event. Kwon and Kang (2016) used time series location map to compare the severity of floods from Twitter posts. Tweets were classified using a 5-by-5 risk evaluation matrix and displayed as color-coded points to assess risk severity level. Onorati et al. (2019) used different techniques, including tree map, word cloud, bubble chart, and an animated map to better analyze information for decision-makers. GeoViewer (Tsou et al. 2017) is a real-time tool that provides an interactive display of multi-media content in addition to a map. These techniques are useful to analyze information visually, targeting specific information need from a specific social media channel. However, for effective disaster response and management, a tool shall provide online access to a variety of useful information, to both citizens and responders in a crisis. To this end, automatic capabilities are required to integrate and fuse information and provide analysis, querying, triaging, and filtering functionality. In this project, we are developing capabilities to effectively disseminate information to the citizens and emergency response worker coordinating disaster related activities on the ground. A sample of some screenshots similar to those we intend to develop is shown in Fig. 5**.** The system will provide the following functionalities.

Situational Awareness Dashboard: the dashboard is concerned with processed, integrated, multimodal information that can be used for decision making. It will allow real-time spatiotemporal distribution of incident and response reports. The information will be visualized using heat maps, time plots, incidents graphs showing a relationship between events, and statistics about the different incidents. With this

[11] https://data.humdata.org/organization/hdx

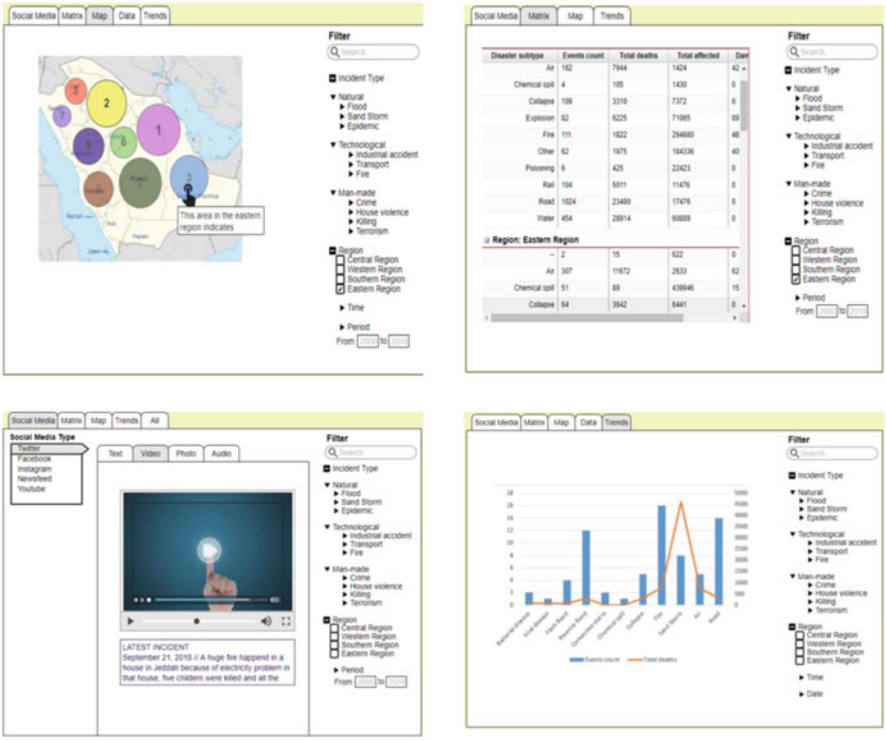


Fig. 5 Sample dashboard visualization of social media analytics framework for crisis management

kind of visual analysis, the emergency response operators will understand how a situation unfolds and help them in disaster planning.

Query and Navigation Interfaces: Intuitive query and navigation interfaces will be developed to allow users to filter through the detail of incidents and events**.** This will allow them to get additional detail on each incident type (such as location, damage, etc.).

Selected Dissemination of Information (SDI): In order to effectively share the right information with the right people, an SDI component will be developed. Such an efficient filtering system will identify only the right content to be shared out of a huge volume of information. For example, if the extracted knowledge is about a traffic accident, the traffic police will be notified with summarized information about the location, the number of injuries and deaths, the services needed etc. Similarly, the information related to medical emergencies will be shared with emergency services such as ambulance services and so on.

3 Discussion

The vast amount of information that is growing by the seconds in the Internet and various social media platforms is enhancing the relevance of information derived from such sources to operational crisis management. One of the key challenges is the availability of suitable computer-based solutions for automated information extraction and analysis from all types of data sources (text, images, audio-video and geo-spatial) and seamless flow of information with relevant stakeholders working together in emergency and crisis management. In the last few years, there have been several interesting developments in developing relevant solutions, their uses are still mostly restricted to a specific information need and single users within an organization and are not yet widely used in operational crisis management. The extension relevant applications to include operational crisis management would contribute towards stimulating the research community to make further advances in related research covering social media, and high-performance computing technologies to process all different type user generated content.

The proposed disaster social media framework described here has several implications for practice and research. From a practical perspective, it presents the robust techniques that can be used to facilitate the development of tools and implementation process to support seamless access to useful, actionable information in a timely fashion for effective disaster operations and decision making in a smart city. The key capabilities are analysis and visualization of multi-model data from diverse sources and customization according to different stakeholders' needs. For instance, is it a group interested in a rescue operation? Is it an organization that provides food or medicines? However, extracting different types of information sought by different stakeholders from millions of diverse sources and modalities is challenging. First, there are no tools available that support the seamless integration of actionable information; therefore, much of the data is underutilized. Second, extracting the conversation structure in social media content and locating actionable information for human interpretation to support multi-dimensional functionalities requires multi-media data processing. Third, much of the information is human-generated; while such information has the advantage of effective disaster response, it could also suffer from false information. Automation is needed to identify the misinformation and integrate the correct information so that human stakeholders can act on such information accordingly.

The technologies we described in all the tiers of the proposed framework, such as event extraction, semantic analysis for multi-media data, event management using disaster ontology, and visualization, aim to address the problem of converting available data into actionable information. The techniques described under each tier provide an opportunity to integrate social media intelligence into disaster operations. While significant work is needed to fully integrate these emerging techniques into a complex disaster system, individuals and society's possible benefit justifies the investment.

Funding Statement The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia, for funding this research work through project number 523.

Acknowledgments The authors extend their appreciation to the Deanship of Scientific Research (DSR) and College of Computer Science and Information Technology at King Faisal University, Saudi Arabia, for facilitating this research work.

References

Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., & Tao, K. (2012). Twitcident: fighting fire with information from social web streams. Proceedings of the 21st International Conference on World Wide Web, New York, United States.

Alam, F., Ofli, F., & Imran, M. (2018a). Crisismmd: Multimodal twitter datasets from natural disasters. Twelfth International AAAI Conference on Web and Social Media, California, USA.

Alam, F., Ofli, F., & Imran, M. (2018b). Processing social media images by combining human and machine computing during crises. *International Journal of Human–Computer Interaction, 34*(4), 311-327.

Angelidou, M. (2015). Smart cities: A conjuncture of four forces. *Cities, 47*, 95-106.

Barz, B., Schröter, K., Münch, M., Yang, B., Unger, A., Dransch, D., & Denzler, J. (2019). Enhancing Flood Impact Analysis using Interactive Retrieval of Social Media Images. *arXiv preprint arXiv:1908.03361*.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*.

Castillo, C. (2016). *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press.

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., & Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics, 38*, 51-60.

Chino, D. Y., Avalhais, L. P., Rodrigues, J. F., & Traina, A. J. (2015). Bowfire: detection of fire in still images by integrating pixel color and texture analysis. 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, Bahia, Brazil.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A., & Scholl, H. J. (2012). Understanding smart cities: An integrative framework. 45th Hawaii international conference on system sciences, Hawaii USA.

Copernicus. (2018). *The European Earth Observation Programme Emergency Management Service*. http://www.copernicus.eu/main/overview

Cross, R. (2010). *Emergency Social Media Summit*. https://redcrosschat.org/about-the-emergency-social-data-summit/

Cunningham, H. (2002). GATE: A framework and graphical development environment for robust NLP tools and applications. Proc. 40th annual meeting of the association for computational linguistics (ACL 2002), Philadelphia.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. IEEE computer society conference on computer vision and pattern recognition (CVPR'05), San Diego, CA, USA.

Dao, M.-S., Quang Nhat Minh, P., Kasem, A., & Haja Nazmudeen, M. S. (2018). A Context-Aware Late-Fusion Approach for Disaster Image Retrieval from Social Media. Proceedings of ACM on International Conference on Multimedia Retrieval, Yokohama, Japan.

Dell'Acqua, F., & Gamba, P. (2012). Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE, 100*(10), 2876-2890.

Dhingra, M., & Chattopadhyay, S. (2016). Advancing smartness of traditional settlements-case analysis of Indian and Arab old cities. *International Journal of Sustainable Built Environment, 5*(2), 549-563.

Dwibedi, D., Malisiewicz, T., Badrinarayanan, V., & Rabinovich, A. (2016). Deep cuboid detection: Beyond 2d bounding boxes. *arXiv preprint arXiv:1611.10010*.

Erhan, D., Szegedy, C., Toshev, A., & Anguelov, D. (2014). Scalable object detection using deep neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, Ohio.

Erk, K., & Pado, S. (2006). Shalmaneser-a flexible toolbox for semantic role assignment. Proceedings of LREC, Genoa, Italy.

Fan, C., Wu, F., & Mostafavi, A. (2020). A Hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. *IEEE Access, 8*, 10478-10490.

Farhadi, A., & Redmon, J. (2018). Yolov3: An incremental improvement. *Computer Vision and Pattern Recognition,, 1804.02767*.

Finkel, J. R., Grenager, T., & Manning, C. D. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), Michigan.

Francalanci, C., Guglielmino, P., Montalcini, M., Scalia, G., & Pernici, B. (2017). IMEXT: A method and system to extract geolocated images from Tweets—Analysis of a case study. 11th International Conference on Research Challenges in Information Science (RCIS), Brighton, United Kingdom.

Gaur, M., Shekarpour, S., Gyrard, A., & Sheth, A. (2019). empathi: An ontology for emergency managing and planning about hazard crisis. IEEE 13th International Conference on Semantic Computing (ICSC), Newport Beach, CA, USA.

Girshick, R. (2015). Fast r-cnn. Proceedings of the IEEE international conference on computer vision, Cambridge, MA, USA.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. Proceedings of the IEEE international conference on computer vision, Cambridge, MA, USA.

Huang, X., Wang, C., Li, Z., & Ning, H. (2019). A visual–textual fused approach to automated tagging of flood-related tweets during a flood event. *International Journal of Digital Earth, 12*(11), 1248-1264.

Huyck, C. K., & Adams, B. J. (2002). *Emergency response in the wake of the World Trade Center attack: The remote sensing perspective* (Vol. 3). MCEER Buffalo, New York.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013). Extracting information nuggets from disaster-Related messages in social media. Iscram, Germany.

Ionescu, B., Radu, A.-L., Menéndez, M., Müller, H., Popescu, A., & Loni, B. (2014). Div400: a social image retrieval result diversification dataset. Proceedings of the 5th ACM Multimedia Systems Conference, Singapore.

Khatoon, S., Romman, L. A., & Hasan, M. M. (2020). Domain Independent Automatic Labeling system for Large-scale Social Data using Lexicon and Web-based Augmentation. *Information Technology and Control, 49*(1), 36-54.

Kireyev, K., Palen, L., & Anderson, K. (2009). Applications of topics models to analysis of disaster-related twitter data. NIPS workshop on applications for topic models: text and beyond, Whistler, Canada.

Klein, D., & Manning, C. D. (2002). Fast exact inference with a factored model for natural language parsing. *Advances in neural information processing systems, 15*, 3-10.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*, 1097-1105.

Kumar, S., Barbier, G., Abbasi, M. A., & Liu, H. (2011). Tweettracker: An analysis tool for humanitarian and disaster relief. Fifth international AAAI conference on weblogs and social media, Barcelona, Spain.

Kwon, H. Y., & Kang, Y. O. (2016). Risk analysis and visualization for detecting signs of flood disaster in Twitter. *Spatial information research, 24*(2), 127-139.

Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008). Learning realistic human actions from movies. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. European conference on computer vision, Zurich, Switzerland.

Liu, X., Wei, F., Zhang, S., & Zhou, M. (2013). Named entity recognition for tweets. *ACM Transactions on Intelligent Systems and Technology (TIST), 4*(1), 1-15.

Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. International Symposium on Music Information Retrieval (ISMIR), Plymouth, Massachusetts.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., & Blanford, J. (2011). Senseplace2: Geotwitter analytics support for situational awareness. IEEE conference on visual analytics science and technology (VAST), Providence, RI, USA.

Malik, M. M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). Population bias in geotagged tweets. Ninth international AAAI conference on web and social media, Oxford, England.

Mason, D. C., Speck, R., Devereux, B., Schumann, G. J.-P., Neal, J. C., & Bates, P. D. (2010). Flood detection in urban areas using TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing, 48*(2), 882-894.

Meier, P. (2015). *Digital humanitarians: how big data is changing the face of humanitarian response*. Routledge.

Middleton, S. E., Middleton, L., & Modafferi, S. (2013). Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems, 29*(2), 9-17.

Mishler, A., Crabb, E. S., Paletz, S., Hefright, B., & Golonka, E. (2015). Using structural topic modeling to detect events and cluster Twitter users in the Ukrainian crisis. International Conference on Human-Computer Interaction, California, USA.

Murthy, D., Gross, A., & McGarry, M. (2016). Visual social media and big data. Interpreting Instagram images posted on Twitter. *Digital Culture & Society, 2*(2), 113-134.

Najibi, M., Rastegari, M., & Davis, L. S. (2016). G-cnn: an iterative grid based object detector. Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, Nevada.

Nguyen, D. T., Alam, F., Ofli, F., & Imran, M. (2017a) *Automatic image filtering on social networks using deep learning and perceptual hashing during crises* Proceedings of the 14th International Conference on Information Systems for Crisis Response And Management (ISCRAM), Albi, France.

Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017b). Damage assessment from social media imagery data during disasters. Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Sydney Australia.

Okolloh, O. (2009). Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action, 59*(1), 65-70.

Olteanu, A., Vieweg, S., & Castillo, C. (2015). What to expect when the unexpected happens: Social media communications across crises. Proceedings of the 18th ACM conference on computer supported cooperative work & social computing, Vancouver BC Canada.

Onorati, T., Díaz, P., & Carrion, B. (2019). From social networks to emergency operation centers: A semantic visualization approach. *Future Generation Computer Systems, 95*, 829-840.

Peters, R., & de Albuquerque, J. P. (2015). Investigating images as indicators for relevant social media messages in disaster management. ISCRAM, Kristiansand, Norway.

Purkait, P., Zhao, C., & Zach, C. (2017). SPP-Net: Deep absolute pose regression with synthetic views. *arXiv preprint arXiv:1712.03452*.

Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Mining crisis information: A strategic approach for detection of people at risk through social media analysis. *International journal of disaster risk reduction, 27*, 556-566.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, San Juan, PR, USA.

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence, 39*(6), 1137-1149.

Ritter, A., Clark, S., & Etzioni, O. (2011). Named entity recognition in tweets: an experimental study. Proceedings of the 2011 conference on empirical methods in natural language processing, Edinburgh, Scotland, UK.

Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., & Laredo, J. A. (2013). CrisisTracker: Crowdsourced social media curation for disaster awareness. *IBM Journal of Research and Development, 57*(5), 4: 1-4: 13.

Saied, A. (2020). *FIRE Dataset: Outdoor-fire images and non-fire images for computer vision tasks.*https://www.kaggle.com/phylake1337/fire-dataset

Schnebele, E., & Cervone, G. (2013). Improving remote sensing flood assessment using volunteered geographical data.

Seppanen, H., & Virrantaus, K. (2015). Shared situational awareness and information quality in disaster management. *Safety Science, 77*, 112-122.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Starbird, K., & Stamberger, J. (2010). Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. Proceedings of the 7th International ISCRAM Conference, Seattle, WA.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, Massachusetts.

Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.

Tsou, M.-H., Jung, C.-T., Allen, C., Yang, J.-A., Han, S. Y., Spitzberg, B. H., & Dozier, J. (2017). Building a real-time geo-targeted event observation (Geo) viewer for disaster management and situation awareness. International cartographic conference, Springer, Cham.

Wang, H., & Schmid, C. (2013). Action recognition with improved trajectories. Proceedings of the IEEE international conference on computer vision, Darling Harbour, Sydney.

Yoo, D., Park, S., Lee, J.-Y., Paek, A. S., & So Kweon, I. (2015). Attentionnet: Aggregating weak directions for accurate object detection. Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile.

Yu, M., Yang, C., & Li, Y. (2018). Big Data in Natural Disaster Management: A Review. *Geosciences, 8*(5), 165.

Zhang, C., Fan, C., Yao, W., Hu, X., & Mostafavi, A. (2019). Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of Information Management, 49*, 190-207.

Zhao, Z.-Q., Zheng, P., Xu, S.-t., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems, 30*(11), 3212-3232.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy, 2*(2), 7-33.