This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2939496, IEEE Internet of Things Journal

A Synergetic Orchestration of Objects, Data and Services to Enable Smart Cities

Linlin You, *Member, IEEE,* Bige Tunçer, *Member, IEEE,* Rui Zhu, Hexu Xing, Chau Yuen, *Senior Member, IEEE,*

***Abstract*—Smart Cities (SC), as a novel solution built on top of large scale IoT systems, experiences a rapid growth worldwide, in which, a synergetic orchestration among objects, data, and services is emphasized for innovative solutions to elevate the intelligence of cities based on the fusion of multi-source and multi-modal data. To enable such orchestration, this paper proposes a Smart Service Orchestration Architecture (SSOA) to coordinate ubiquitous objects, create interlinked data, and implement versatile smart services. As a proof of concept of SSOA, an Informed Design Platform (IDP) is presented to demonstrate how a smart service system can be designed and how a synergetic orchestration can be implemented to support an informed place design. Moreover, two dedicated mechanisms, namely Place Utilization Analysis Mechanism (PUAM) and Ensemble-based Activity Detection Mechanism (EADM), are implemented in a multi-source data processing flow to illustrate how massive geo-referenced data can be analyzed effectively and efficiently by machine learning algorithms to extract key information for a comprehensive data fusion required in using multi-modal IoT systems. As evaluated, PUAM running in a distributed environment can dramatically improve the performance of geospatial clustering about 11 times from the baseline 53.6s to 4.7s, and EADM with an ensemble activity classifier achieves the highest accuracy about 87.7% and also the highest f-score per activity category. Finally, various insights about the project testbed Jurong East, Singapore are discussed to reveal its place design context.**

***Index Terms*—Smart Cities, Synergetic Orchestration, Multi-source Data Processing, Informed Design, Place Utilization and Activity Analysis, Social Media Data Analysis, Geospatial Clustering, Ensemble Learning**

I. Introduction

SMART Cities (SC) as a ICT (Information and Communications Technology) solution is innovated to embrace a digital transformation of cities in supporting the efficiency of governance, ensuring the competitiveness of business, and improving the quality of life by addressing emerging and critical issues, such as the operational isolation in multi-modal IoT systems, the unprecedented explosion of big data from multiple sources, and the rapid growth of demands for more intelligent services [1]–[3]. In general, the key challenge in enabling such solution is about how to harness advanced technologies, e.g., IoT, Artificial Intelligence (AI), Big Data, Cloud/Fog Computing and other prevailing ones, to support an orchestration of three SC elements, namely, 1) "Objects" incorporating IoT [4] with virtual systems (e.g., social networks, etc.), 2) "Data" integrating massive and heterogeneous data from multiple sources [5], and 3) "Services" providing customizable value propositions [6].

Currently, many initiatives implement such orchestration to elevate the intelligence of cities by integrating a wide varieties of sensors/IoT devices to gather more comprehensive data, which are further processed, fused, mined and utilized to support diverse demands in various SC domains, e.g., in mobility to provide on-demand and personalized mobility services [7], in environment to solve pollution issues [8], in governance to create a collaborative city management process [9], in health-care to support personal e-health [10], and in security to improve safety and emergency management [11]. Specifically, it becomes critical about how to enable multi-model sensor fusion for the collection of multi-source data with spatiotemporal, contextual, and personalized information, e.g., place centric time series data obtained by environmental sensors, people centric trajectory data obtained by smartphones or Wi-Fi sniffers, and user preference data obtained from social media or user stated/revealed preference surveys.

In order to overcome the challenge and further impel the development of SC, a synergetic orchestration architecture is required that enables multi-source and multi-modal information to be integrated in a single framework. Such design is important for SC systems/services built on top of large scale IoT systems, as, in general, they consist of sensors of multiple modality and, therefore, generate data with significant heterogeneity. As an important topic in the field of IoT, related architectures and systems are discussed with a shift from the initial emphasis on the accessibility of objects and services [12], [13] to the the succedent design of standardized systems reusing infrastructure, data and services [6], [14]–[16].

However, current solutions only support a part of the whole orchestration process, in which, objects, data and services shall be coordinated comprehensively, by tearing down 1) vertical orchestration barriers about how to adaptively collect and uniformly manage data generated from objects and used by services; and 2) horizontal orchestration barriers about how to collaborate various object networks for their full potentials, how to create an interlinked and extensible data network by addressing issues related to big and heterogeneous multi-source data processing, and how to modularize and reuse micro-services (service modules) for diverse value propositions.

L. You is with Intelligent Transportation Systems Lab, Massachusetts Institute of Technology, MA, USA, 02139.

B. Tunçer, and H. Xing are with the Architecture and Sustainable Design, Singapore University of Technology and Design, Singapore, 487372.

R. Zhu is with the Future Urban Mobility IRG, Singapore-MIT Alliance for Research and Technology Centre, Singapore, 138602.

Chau Yuen is with the Engineering Product and Design, Singapore University of Technology and Design, Singapore, 487372.

E-mail: linlin@smart.mit.edu and bige_tuncer@sutd.edu.sg

In order to address these issues, this paper proposes a Smart Service Orchestration Architecture (SSOA) to unify the design of service systems based on a three-tier architecture consisting of Cloud of Objects (COO), Cloud of Data (COD) and Cloud of Services (COS), and to support the synergetic orchestration among objects, data and services by a multi-source data processing flow, which can coordinate components in the three tiers to collect data from multiple objects (such as large scale IoT systems and user systems) adaptively, create interlinked data network effectively and efficiently, and build value proposition agilely. As a proof of concept, an Informed Design Platform (IDP) developed in a "Livable Places" project [17] is presented to illustrate how SSOA segregates system components of IDP into the three tiers, and then incorporates them for a place utilization and activity analysis flow, which translates massive geo-referenced data into place design insights with the support of two machine learning mechanisms, namely Place Utilization Analysis Mechanism (PUAM) and Ensemble-based Activity Detection Mechanism (EADM).

The remainder of this paper is structured as follows. First, section II introduces SC orchestration with the definition of three orchestration elements, the summary of orchestration barriers and the review of related solutions. Second, SSOA is proposed in section III, and IDP together with a place and utilization analysis flow is discussed in section IV. Third, section V evaluates the performance of PUAM and EADM, and then discusses the observations of the project testbed Jurong East, Singapore and also the key learnings from SSOA. Finally, section VI concludes the work and sketches the future.

II. The introduction of SC orchestration

This section introduces SC orchestration by defining three orchestration elements, summarizing two kinds of orchestration barriers and comparing related orchestration solutions.

A. Three orchestration elements

As shown in Figure 1, SC coordinates three key elements, namely objects, data and services in various SC domains:

- **SC Domain**: It defines the scope of a solution, such as a) Smart Mobility to optimize the demand and supply of mobility [7]; b) Smart Governance to innovate city governance process [9]; c) Smart Environment to build an eco-friendly and sustainable city [8]; d) Smart Living to enhance living utilities and experience [10], [11].
- **Objects**: More scaled than IoT [2], [4], [14], [18], they are objects embedded in the city and used by the public, including ubiquitous physical objects (e.g., sensors, smartphones, etc.) and collaborative virtual objects (e.g., social networks, online systems, etc.)
- **Data**: It is represented by an interlinked and extensible data network to manage data generated from multiple "Objects" with 4V characteristics, namely big Volume, large Variety, high Velocity, and diverse Value [2], [14].
- **Services**: They implement innovative processes and sophisticated analytics to support the efficiency, livability and sustainability of the city by coordinating reusable and modularized service components [6], [17].

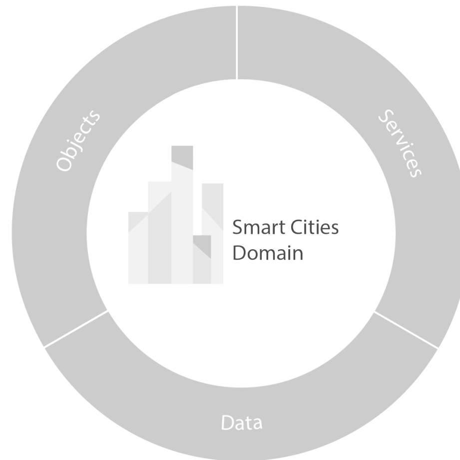


Fig. 1. The elements of SC orchestration

In summary, SC elevates the intelligence of various domains by a) interconnecting infrastructure, people, systems and every connectable object, b) collecting, fusing, storing, exchanging and mining big and heterogeneous data, and c) implementing various smart services for a livable and sustainable city.

B. Orchestration barriers (OB)

As shown in Figure 2, several vertical and horizontal barriers shall be addressed to enable a synergetic orchestration.

1) Vertical orchestration barriers among three elements: A vertical collaboration among three elements is illustrated by two interactions between "Objects" and "Data" and between "Data" and "Services" to support data collection, fusion, mining and visualization [4], [16]. In general, there are two issues:

- OB1: How to simplify the data collection process. An adaptive data collection mechanism is needed to incorporate various data access methods of "Objects" [13], [17].
- OB2: How to unify the data management process. Standardized methods are required to support data CRUD (Create, Read, Update, Delete) operations [12], [19].

2) Horizontal orchestration barriers in objects: Due to the technical difference, intrinsic boundaries exist among object networks, e.g., large scale IoT networks, social networks, etc., and, therefore, prevent the orchestration. In general, there are two challenges:

- OB3: How to intertwine various object networks. The ultimate output "data" become important for a SC solution to mediate differences among object networks used. Hence, a common data access interface is required to replace ad-hoc methods, e.g., database connections, data retrieval APIs, shared files, etc., for the integration [17].
- OB4: How to explore the full potential of object networks. Even though an object network can support multiple tasks, e.g., a Wi-Fi network can be used as a people sniffer or an indoor navigation facility [18], [20], its data generated are consistent. Therefore, the value of object networks can be maximized by fully utilizing their data.

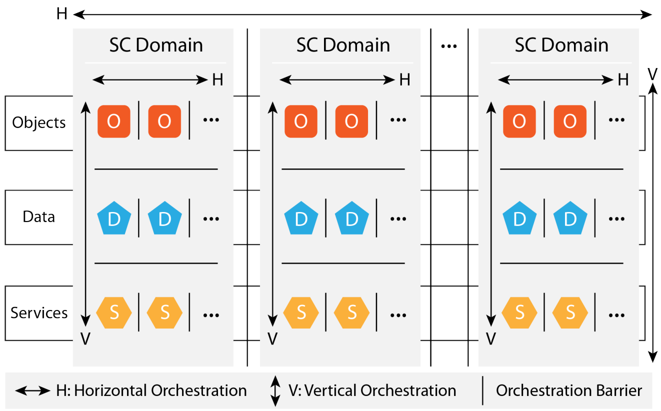


Fig. 2. An abstracted diagram of synergetic orchestration

3) Horizontal orchestration barriers in data: The horizontal orchestration of data implies a unified and scalable data integration and storage solution to solve two critical issues:

- OB5: How to create an interlinked and interoperable data network. The variety and complexity of multi-source data shall be eased through a heterogeneous data integration mechanism, which can extract key information from overwhelming data and fuse them for an interlinked and interoperable data network [17], [21].
- OB6: How to process and store massive data efficiently and effectively. Big data frosts the conventional data processing and storage solutions. Therefore, new types of technologies, such as distributed computing like cloud and fog, shall be adopted to fulfill related needs [6].

4) Horizontal orchestration barriers in services: The reusability of services propels a solution-level redeployment and a microservice-level reuse [10], [15]. Even though the first mode has been widely adopted, the second mode shall be advocated to foster the service innovation. Such that, a general mechanism to modularize and integrate smart services is required, which shall solve two issues [6], [14]:

- OB7: How to modularize services. Service modularization requires a rational category and standard for third parties to define, publish and discover reusable service modules (SMs) more efficiently and effectively.
- OB8: How to integrate service components. A loosely coupling integration mechanism shall be studied to integrate reusable service modules, which shall be invoked and deployed based on actual usage and performance.

C. Related orchestration solutions

In order to tackle aforementioned orchestration barriers, some solutions are proposed in recent years:

- Service Access Orchestration Model (SAOM) (2015) [12]: A mechanism to support the discovery, recruitment, orchestration and billing of cloud-based smart services through a centralized service orchestration unit.
- App Execution Platform (AEP)(2015) [13]: A platform to design, deploy and execute IoT applications by interacting with smart objects through composite application

TABLE I
THE OVERALL EVALUATION OF REVIEWED SOLUTIONS
(●:SUPPORTED; ◐:PARTIALLY SUPPORTED; ○:NOT SUPPORTED)

Solutions	Vertical		Horizontal					
	OB1	OB2	OB3	OB4	OB5	OB6	OB7	OB8
SAOM (2015)	◐	○	◐	○	○	◐	●	●
AEP (2015)	◐	○	●	●	○	◐	○	◐
SCOS (2016)	◐	●	◐	●	○	●	●	●
FogFlow (2017)	●	◐	●	●	○	●	◐	●
OrganiCity (2018)	◐	◐	●	●	◐	●	◐	●
SSOA (2019)	●	●	●	●	●	●	●	●

Orchestration Barrier (OB):
OB1: Data Collection; OB2: Data Management; OB3: Object Integration; OB4: Object Reuse; OB5: Common Integrated Data Model; OB6: Massive Data Processing and Storage; OB7: Service Component Modularization; OB8: Service Module Integration

- Smart City Operating System (SCOS) (2016) [14], [15]: a cloud-enabled ecosystem with infrastructure, data and application layers to create applications based on the scale of a city, and the needs of stakeholders
- FogFlow (2017) [6]: A platform based on distributed computing, such as cloud and fog, to integrate with geo-distributed infrastructure resources and process big data through standardized interfaces.
- OrganiCity (2018) [16]: An open city environment managing objects and data uniformly, and providing common and well-established APIs to upper services, such as city dashboard and citizen service and applications.
- Smart Service Orchestration Architecture (2019): A three-tier service architecture to guide the design and implementation of smart services and to support the synergetic orchestration among objects, data and services.

As shown in Table I, these solutions are evaluated based on their abilities to tear down the vertical and horizontal orchestration barriers defined. In summary, AEP emphasizes the integration and reuse of objects (mainly IoT systems), and SAOM focuses on the horizontal orchestration of services. In common, both of them are lack of mechanisms to address the vertical interaction between data and services, and the fusion of multi-source data. Furthermore, both SCOS and FogFlow can support the two vertical interactions among the three elements, and the three horizontal orchestrations. However, they miss a common data model to harness heterogeneous multi-source data. One step ahead, OrganiCity can address all defined issues, but it is still weak in supporting the two vertical orchestrations, the fusion of multi-source data, and the modularization of service components. Finally, as a standardized architecture, SSOA overcomes the above solutions with the ability to support both horizontal and vertical orchestrations.

III. SSOA: SMART SERVICE ORCHESTRATION ARCHITECTURE

As shown in Figure 3, Smart Service Orchestration Architecture (SSOA) consists of Cloud Of Objects (COO) to

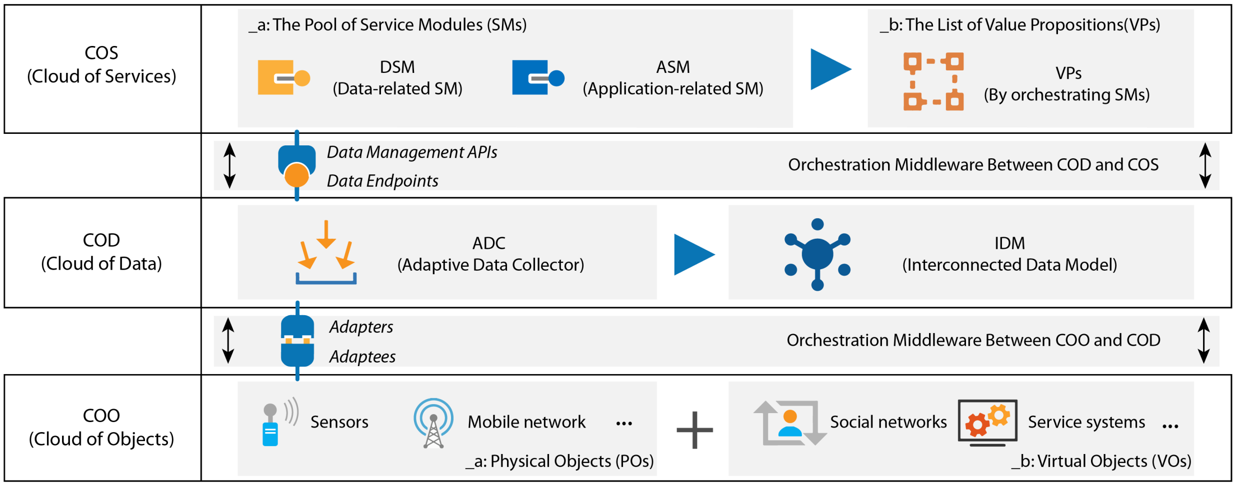


Fig. 3. The three-tier abstract structure of SSOA

TABLE II
KEY ACRONYMS INTRODUCED IN SSOA

Acronyms of SSOA (Smart Service Orchestration Architecture) Tiers		
1	COO	Cloud of Objects
2	COD	Cloud of Data
3	COS	Cloud of Services
Acronyms in COO (Cloud of Objects)		
4	PO	Physical Object
5	VO	Virtual Object
Acronyms in COD (Cloud of Data)		
6	ADC	Adaptive Data Collector
7	IDM	Interconnected Data Model
8	PDM	Plain Data Model
9	LDM	Linked Data Model
10	PD	Plain Data
11	LD	Linked Data
12	CI	Common Information
13	DI	Domain Information
14	AD	Analysis Dimension
15	AM	Analysis Measure
Acronyms in COS (Cloud of Services)		
16	SM	Service Module
17	DSM	Data-related Service Module
18	ASM	Application-related Service Module

manage diverse physical and virtual objects, Cloud Of Data (COD) to collect, store and distribute multi-source data, and Cloud Of Services (COS) to modularize services into reusable components for value propositions. Additionally, SSOA also enables a multi-source data processing flow supporting a synergetic orchestration among objects, data and services.

For the sake of presentation and readability, Table II lists key acronyms used in SSOA.

A. Cloud of Objects (COO)

In a digitalized city, objects can be connected via networks to form "Cloud of Objects" (COO), which includes not only physical objects (POs), such as sensors, mobile phones, and alike IoT devices, but also virtual objects (VOs), such as social networks, websites and alike user systems. In order to alleviate fundamental variations in COO, SSOA designs a universal object management mechanism to ensure a seamless exchange of data generated by objects via a pair of "Adaptee" and "Adapter", which encapsulates an "Adaptee" (which provides ad-hoc data access methods) of an object by an "Adapter" with a common data extraction interface. Such that, the interaction between COO and COD can be simplified, and the extensibility of COO can be ensured, as a new object can be easily incorporated through its pair of "Adaptee" and "Adapter".

B. Cloud of Data (COD)

It includes three components, namely Adaptive Data Collector (ADC), Interconnected Data Model (IDM) and Data Endpoints (DEPs), to support multi-source data collection, integration, storage, and distribution.

1) Adaptive Data Collector (ADC): ADC defines a list with pairs of "Adaptee" and "Adapter" to implement a common data collection process with high scalability. Based on the list, ADC can also be easily extended and degraded to collect data from multiple sources continuously or periodically.

2) Interconnected Data Model (IDM): As shown in Figure 4 (A), IDM manages heterogeneous multi-source data in an interlinked and interoperable data network with two models, i.e., the plain data model (PDM) and linked data model (LDM). Specifically, PDM categorizes the information from each object into a) common information (CI) that is shared among objects, e.g., temporal and spatial information, and b) domain information (DI) that can only be properly interpreted with domain knowledge, e.g., the reading of motion sensors. Correspondingly, PDM stores plain data (PD) as defined in formula 1, where i indicates the $i-th$ object, and n is the total number of objects in COO; CI is a set of common information, and its total number is N_{CI}; DI_i contains a set of values V defined in the $i-th$ object, and its total number is N_{DI_i}.

$$PD_i = \{CI, DI_i\}$$
$$s.t. \begin{cases} i = 1\ to\ n \\ CI = \{CI_j \mid j = 1\ to\ N_{CI}\} \\ CI_j = \begin{cases} value, & \text{if } CI_j \text{ exists} \\ 0, & \text{otherwise} \end{cases} \\ DI_i = \{V_k \mid k = 1\ to\ N_{DI_i}\} \end{cases} \tag{1}$$

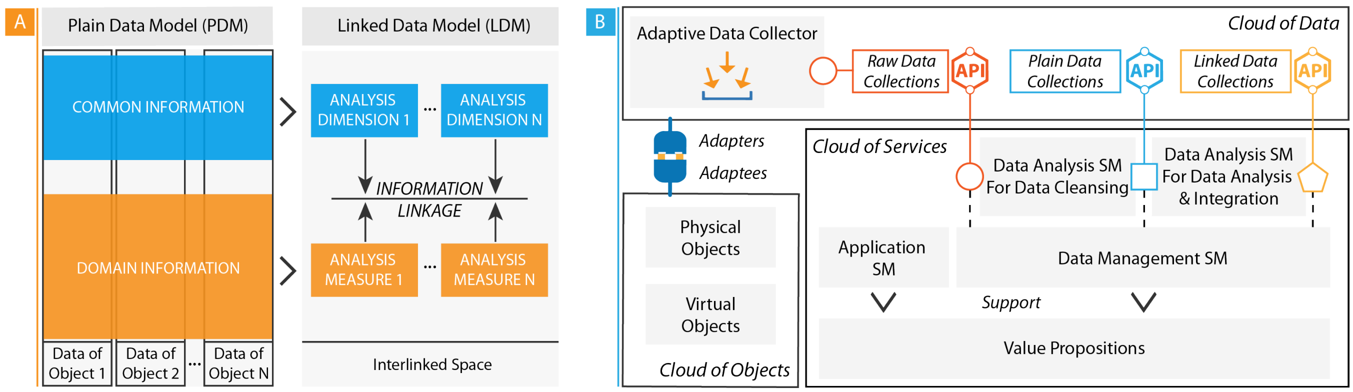


Fig. 4. (A) Interconnected data model of SSOA: (A.1) Plain Data Model (PDM), and (A.2) Linked Data Model (LDM); (B) Abstracted multi-source data processing flow of SSOA (running in a distributed environment)

Moreover, LDM consists of a) analysis dimension (AD) that is generated from CI with normalized attributes, e.g., the temporal CI can be transformed into an AD with attributes such as year, month, day, hour, and minute; b) analysis measure (AM) that stores the processed DI with properties about source, value, and measurement, e.g., "source: temperature sensor, value: 30, measurement: degree"; and c) information linkage (IL) that specifies the relationship between ADs and AMs, and maintains the scalability of model. Correspondingly, it manages linked data (LD) as defined in formula 2, where i indicates the $i-th$ AMs and its total number is N_{AM}; AD_i is a set of ADs linked with AM_i, and its total number is N_{AD}; IL_i is a vector defining the presence of ADs that are linked with AM_i.

$$
\begin{aligned}
& LD_i = \{IL_i,\ AD_i,\ AM_i\} \\
s.t.\ & \begin{cases} i = 1\ to\ N_{AM} \\ j = 1\ to\ N_{AD} \\ IL_{i,j} = \begin{cases} 0, & \text{if } AD_j \text{ exists} \\ 1, & \text{otherwise} \end{cases} \\ AD_{i,j} = \begin{cases} AD_j, & \text{if } AD_j \text{ exists} \\ 0, & \text{otherwise} \end{cases} \end{cases}
\end{aligned} \tag{2}
$$

3) Data Endpoints (DEPs): DEP is designed as a common portal to manage data in Interconnected Data Model (IDM), and support the vertical orchestration between COD (Cloud of Data) and COS (Cloud of Services). Therefore, standardized data management APIs based on Http(s) methods are implemented as listed in Table III.

TABLE III
THE LIST OF STANDARDIZED APIS

Http(s) Method	Operation	Description
GET	Query	To query a dataset according to specified conditions
POST	Save	To save data in the POST request body to a dataset
PUT	Update	To update data records of a dataset according to update conditions
DELETE	Remove	To remove data records matching with given conditions from a dataset

C. Cloud of Services (COS)

To address issues that well developed and tested services are seldom or hardly reused to reduce service implementation costs, and improve service quality, SSOA establishes a pool of reusable service modules to create value propositions. In general, a service module (SM) is a reusable micro-service/service component that implements a basic feature of a service. SSOA defines two SM groups, namely a) data-related SM (DSM) that manages and analyzes data, e.g., standardized data access APIs and a place utilization analysis component; b) application-related SM (ASM) that consumes and presents data, e.g., analytical dashboard.

D. Multi-source data processing flow

As shown in Figure 4 (B), a novel multi-source data processing flow is enabled by SSOA by orchestrating the components defined in COO (Cloud of Objects), COD (Cloud of Data) and COS (Cloud of Services). First, based on the pair of "Adaptee" and "Adapter", data from COO (Cloud of Objects) is collected in real-time or periodically. Then collected multi-source data is extracted and processed by data-related service modules provided by COS (Cloud of Service) according to the Interconnected Data Model defined in COD (Cloud of Data). Finally, value propositions can be implemented based on data and application service modules. It is worth noting that such flow can be deployed in a distributed environment to address performance issues in processing big data.

In summary, SSOA presents a common framework to design smart service system by cooperating large scale IoT and user-related systems. It has three tiers, through which a synergetic orchestration can be enabled to address issues about IoT and user system integration, multi-source and multi-modal data fusion, service module reuse and value proposition creation.

IV. INFORMED DESIGN PLATFORM (IDP)

As a proof of concept of SSOA, IDP is designed and implemented to harness"big data" of IoT and user systems for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2939496, IEEE Internet of Things Journal

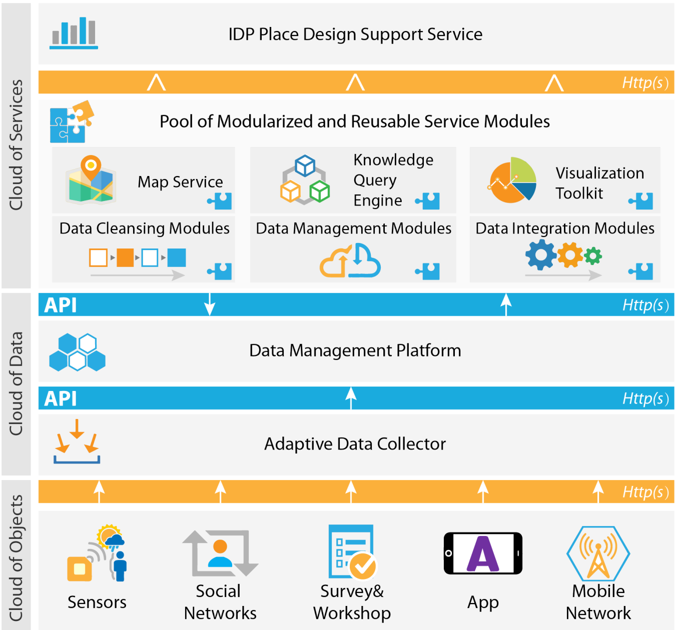


Fig. 5. The abstracted system architecture of IDP

TABLE IV
AN EXAMPLE MMQ TO EXTRACT KNOWLEDGE ABOUT SENTIMENTS

```
SELECT Sum(Sentiment)
WHERE People.ageGroup = ['Elderly']
AND Place.placeName = ['Pavilion']
AND Time.timestamp ≥ T1 AND Time.timestamp ≤ T2
FROM Public Sentiment Linked Data
```

Fig. 6. The user interface of IDP Place Design Support Service

responsive designs increasing the livability of urban spaces. In this section, its system architecture and a multi-source data processing flow with two dedicated analysis mechanisms to analyze place utilization and activity are presented to demonstrate how SSOA supports the synergetic orchestration.

A. IDP System Architecture

As shown in Figure 5, IDP manages its components according to SSOA in three tiers:

1) Components in Cloud of Objects (COO): IDP incorporates 5 objects including a) IoT sensors, which are deployed in the testbed to gather environment and motion data, b) social networks, which are integrated to collect social messages, c) survey and workshop, which are conducted to collect opinions of inhabitants about the project testbed, d) smartphones running with a people-sensing application (App) that collects user reviews about predefined places, and finally, e) a mobile network provided by a local telecom company for cellular data.

2) Components in Cloud of Data (COD): The Adaptive Data Collector (ADC) is implemented to gather data from the aforementioned IoT devices and user-oriented objects by using 5 pairs of "Adaptee" and "Adapters" that encapsulate ad-hoc data access methods, e.g., a database connection of a sensor management system, real-time data streams of social networks, etc. In general, ADC controls adapters to collect data periodically, i.e., daily from sensors, real-time from social networks, weekly from App, on-demand from survey&workshop, and monthly from the mobile network.

Moreover, a Data Management Platform (DMP) is implemented with RESTful APIs to support data CRUD (Create, Read, Update and Delete) operations. Based on the RESTful APIs, components defined in COO (Cloud of Objects) and COS (Cloud of Services) can be integrated loosely. In DMP, an Interconnected Data Model (IDM) is also defined to manage multi-source data with three common Analysis Dimensions (ADs), namely "Place", "Time" and "People", and several Analysis Measures (AMs), e.g., place activity, utilization, sentiment, etc.

3) Components in Cloud of Services (COS): IDP establishes a pool of modularized and reusable service modules (SMs) running in a distributed environment to process multi-source data and create user value propositions. Specifically, as for data processing, a) data management modules provide standardized methods based on RESTful APIs; b) data cleansing modules remove dirty data and reconcile or remove records with missing information; and c) data integration modules extract key information based on machine learning methods, and integrate them by creating analysis dimensions and measures defined in the Interconnected Data Model (IDM).

Moreover, three application-related SMs are implemented, namely 1) a map service to present the testbed map with more details; 2) a knowledge query engine to extract insights from linked data based on a multi-dimension and multi-measure query (MMQ). As shown in Table IV, MMQ can easily extract knowledge about how the "Elderly" feels about the "Pavilion" in the time period from T_1 to T_2; and 3) a visualization toolkit to support user interactions and present insights. By integrating these SMs, an IDP Place Design Support Service is implemented as shown in Figure 6, which supports users to select analysis measures on the left, configure three analysis dimensions on the top, and digest extracted knowledge in analysis charts and on an analysis map in the middle.

In summary, IDP is implemented based on SSOA to coordinate 5 objects, process multi-source data and manage modularized service components for a value proposition to support the informed design. Moreover, it also implements multi-source data processing flows to support the synergetic orchestration and an example will be discussed in the following section.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2939496, IEEE Internet of Things Journal

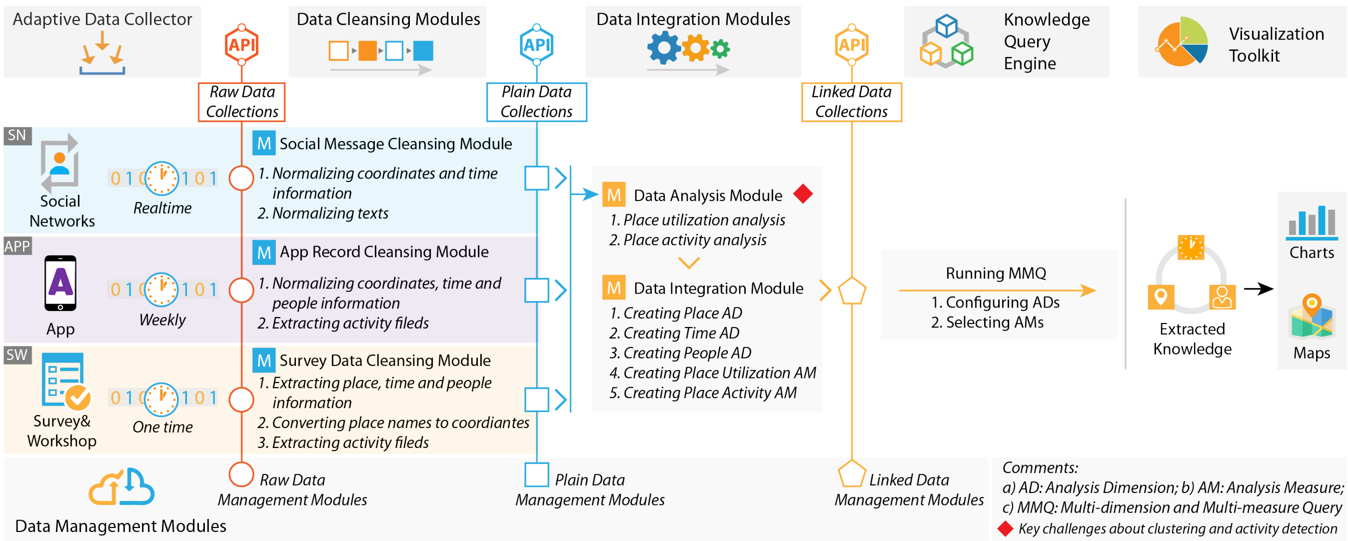


Fig. 7. Multi-source data processing flow to analyze place utilization and activity

B. *Place Utilization and Activity Analysis Flow*

As shown in Figure 7, on top of IoT devices and user systems, a synergetic orchestration among objects, data and services is implemented to reveal place design context by analyzing place utilization and activities. First, the Adaptive Data Collector (ADC) gathers data from social networks (SN) in real-time, smartphones running with the people sensing mobile app (App) per week, and survey&workshop (SW) in one time. Then, collected data are normalized by three data cleansing modules separately to identify three common information, i.e., people, time and place, and extract domain information related to people activities. Specifically, in the social message cleansing module, texts are parsed into cleansed text, hashtags, mentions, and POIs for further analysis. After normalized plain data are ready, they are integrated for linked data by passing through 1) a data analysis module to calculate place utilization indicators, and detect place activities, and 2) a data integration module to create people, time and place analysis dimensions, and place utilization and activity analysis measures. Finally, an IDP Place Design Support Service is created based on 1) the knowledge query engine to extract insights about place utilization and activities from linked data by running multi-dimension and multi-measure queries, and 2) the visualization toolkit to support the configuration of analysis dimensions and the selection of analysis measures, as well as the presentation of extracted knowledge. It is worth noting that all data uploading and downloading flows are supported by corresponding data management modules.

As key mechanisms in the flow to process massive geo-referenced data and social messages to calculate place utilization indicators and detect activities, a Place Utilization Analysis Mechanism (PUAM) and an Ensemble-based Activity Detection Mechanism (EADM) are proposed.

1) Place Utilization Analysis Mechanism: As shown in Figure 8 (A), it first clusters geo-referenced data and then detects frequent places to calculate place utilization indicators.

In Step 1: Clustering, a scalable DBSCAN algorithm is designed with two kinds of computation units, namely, 1) Primary Unit (PU) to partition data, distribute data partitions, and consolidate intermediate clustering results (ICRs) for the final result, and 2) Auxiliary Unit (AU) to run the conventional DBSCAN algorithm on a data partition for an ICR. In order to get N data partitions, PU uses longitude as the division reference, and creates overlapping areas OA_i to link clusters in two adjacent data partitions (noted as P_i and P_{i+1}, $i = 1\ to\ N-1$). In AUs, DBSCAN runs with two parameters configured, i.e., 1) ϵ the distance reachable threshold to decide whether a point shall be added into a group of points or not, and 2) $MinPts$ the minimum points threshold to define whether a group of points can form a cluster or not. Then, PU processes ICRs by marking clusters without points in OA_i as final clusters, and merging clusters with the same points in OA_i to generate final clusters. It is worth noting that $2 \times \epsilon$ is used as the minimal length of OA_i, as if a smaller length is used, some clusters in P_i and P_{i+1} located in OA_i cannot be merged correctly [22].

In Step 2: Frequent Place Detection, an unsupervised frequent place detection algorithm is implemented to detect frequent places by using clusters of 7 consecutive days. First, for each day, candidate frequent places (CFPs) are detected by comparing if a cluster of a reference day overlaps with clusters of rest 6 days. If the overlapping count is bigger than an overlapping cluster number threshold T_V, the cluster together with its overlapping clusters is marked as a CFP. Second, all 7 days' CFPs will be merged if two CFPs contain at least one same cluster to generate a merged frequent place (MFP) with de-duplicated overlapping clusters. Finally, for a MFP, it checks how many distinct days a MFP covers. If the number of distinct days is bigger than a distinct day threshold T_D, the MFP is classified as a frequent place. Based on detected frequent places, place utilization indicators, such as the average number of visitors per day (ANV), the size of the place (SP), and the density of people (DP) are calculated. Specifically, DP is calculated according to formula 3, where $Max(SP)$ gets the maximum value of SP; C is a constant to

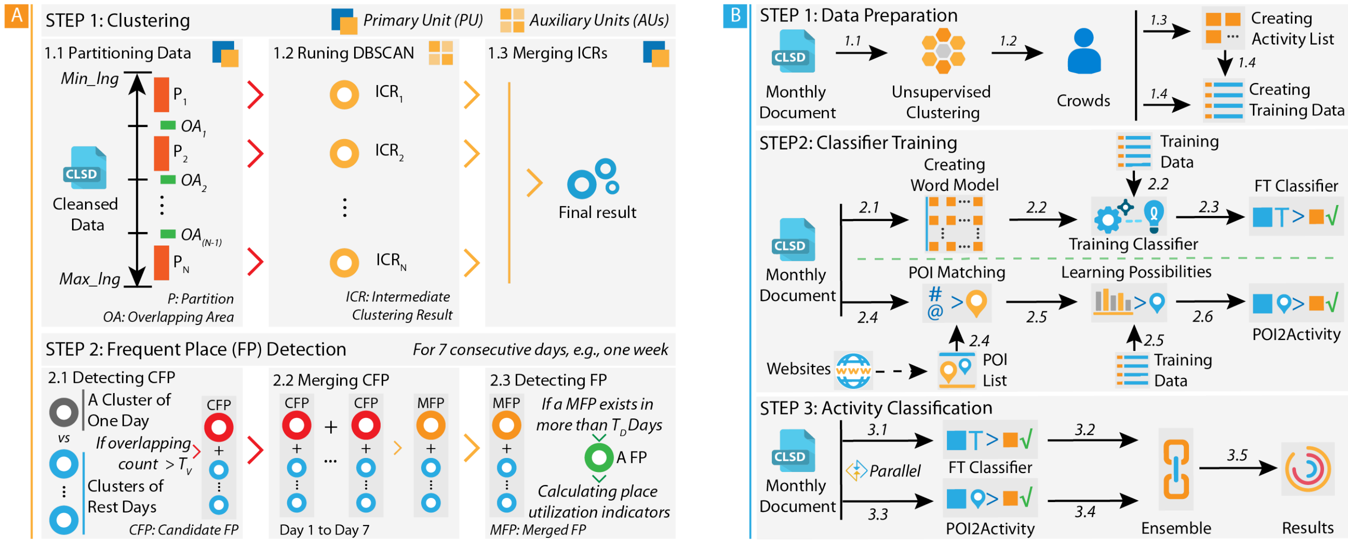


Fig. 8. The summary of two mechanisms: (A) Place Utilization Analysis Mechanism, and (B) Ensemble-based Activity Detection Mechanism

normalize the value, and its default value is 1.

$$DP = \frac{ANV * Max(SP)}{SP * C} \tag{3}$$

Generally, PUAM can accurately calculate place utilization indicators by detecting frequent places with a high clustering performance, which is evaluated in section V-A2.

2) Ensemble-based Activity Detection Mechanism: As shown in Figure 8 (B), it uses monthly documents to detect activities of social messages in three steps:

In Step 1: Data Preparation, crowds are involved to define activity categories and create initial training data. First, unsupervised algorithms are used to create activity clusters, which provide useful information for crowds to better understand the context and accelerate the activity category definition process. Then, about 20% of the records of the monthly document are labeled by crowds to create sufficient training data.

In Step 2: Classifier Training, two classifiers are trained based on a deep learning method and a statistic method respectively. The first method uses a state-of-the-art method FastText [23], which not only can train model in a shorter time compared to other deep learning methods, but also can efficiently handle short texts with char-N-gram modeled. The second method is proposed based on the fact that many social messages contain POI tags, whose functionalities can indicate user activities (e.g., "I am at @Jurong East MRT"). So that, a dedicated procedure is designed to learn a POI2Activity classifier, which contains a possibility vector of a POI indicating which activity is more related. In general, first, a POI list is crawled from related websites, such as shopping mall websites. Then, for the i_{th} POI in the POI list, an initial "POI2Activity" possibility vector IPV_i is created as shown in formula 4, where PA_j is the possibility that the j_{th} activity is associated with the i_{th} POI according to the POI category information crawled.

$$IPV_i = \{PA_j\} \\ s.t. \quad \begin{cases} j \in N \\ \sum_{j=1}^{N} PA_j = 1 \end{cases} \tag{4}$$

Afterwards, a statistical "POI2Activity" possibility vector SPV of a POI is calculated according to formula 5, where TN is the total number of social messages containing the name or aliases of the given POI, AN_j is the number of social messages with the given POI belonging to the j_{th} activity category. If a POI in the POI list does not appear in the training data, a SPV with default zero values is set to it.

$$SPV = \{\frac{AN_j}{TN}\} \\ s.t. \quad \begin{cases} j \in N \\ \sum_{j=1}^{N} AN_j = TN \end{cases} \tag{5}$$

Finally, based on IPV and SPV, a POI2Activity classifier is learned to calculate the final possibility vector PV according to formula 6, where IPV_i is the IPV of the i_{th} POI; SPV_i is the SPV of i_{th} POI; f is a normalization function.

$$PV_i = f(\ \{(IPV_{i_j} + SPV_{i_j}) \mid j = 1\ to\ N\}\ ) \tag{6}$$

In Step 3: Activity Classification, for a social message, results of two classifiers are merged according to an ensemble rule defined in formula 7, where FA is the final detected activity with the highest ensemble score; i is the activity sequence, and N is the total activity number; S_{POI_i} is the score calculated by the POI2Activity classifier; S_{DP_i} is the score generated by the deep learning activity classifier; α and β are weight coefficients derived from empirical experience.

$$FA = \max_{i \in N}(\alpha \times S_{POI_i} + \beta \times S_{DP_i}) \tag{7}$$

In general, EADM measures the syntactic and semantic meaning of texts and also the functional information of mentioned locations to detect public activities based on social messages, and its performance is evaluated in section V-A3.

V. Evaluation and Discussions

In this section, the performance of two analysis mechanisms is evaluated, and the place design context derived from IDP

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2939496, IEEE Internet of Things Journal

A.1

	CUN_1	CUN_3	CUN_6	CUN_9	CUN_12	Improved by CUN
DPN_3	46.0s	45.7s	~~45.7s~~	~~45.7s~~	~~45.7s~~	00.0%
DPN_6	22.9s	19.4s	18.6s	~~18.6s~~	~~18.6s~~	18.8%
DPN_9	19.2s	19.1s	15.4s	15.4s	~~15.4s~~	19.8%
DPN_12	16.2s	15.7s	14.0s	13.9s	12.5s	***22.8%***
Improved by DPN	64.8%	65.6%	69.4%	69.6%	***72.6%***	

SE-DPS: Improvement against baseline 53.6s from 14.2% to 76.7%

A.2

	CUN_1	CUN_3	CUN_6	CUN_9	CUN_12	Improved by CUN
DPN_3	20.3s	9.5s	~~9.5s~~	~~9.5s~~	~~9.5s~~	53.2%
DPN_6	12.5s	6.8s	5.3s	~~5.3s~~	~~5.3s~~	***57.6%***
DPN_9	10.4s	7.6s	5.5s	4.8s	~~4.8s~~	53.8%
DPN_12	10.0s	6.8s	5.2s	5.1s	4.7s	53.0%
Improved by DPN	***50.7%***	28.4%	45.3%	49.5%	50.5%	

NE-DPS: Improvement against baseline 53.6s from 62.1% to 91.2%

Best among CUNs in a DPN *Best among DPNs in a CUN*

B.1 / B.2

M / I	J48	Bayes	RF	SMO	FastText	Sent2Vec	CNN	EADM
ACC.	52.4%	54.8%	70.2%	71.2%	74.5%	75.3%	79.0%	87.6%
TT	797s	21.8s	303s	14.5s	41.6s	475s	1428s	60.0s
A_1	78.7%	80.6%	87.7%	87.5%	91.5%	91.1%	94.2%	95.3%
A_2	57.7%	69.5%	79.6%	79.1%	80.8%	79.1%	87.1%	88.3%
A_3	48.3%	59.7%	73.3%	74.0%	80.8%	79.7%	83.7%	89.6%
A_4	53.6%	53.3%	72.3%	73.3%	68.8%	70.0%	73.2%	91.2%
A_5	78.0%	80.5%	85.0%	84.2%	77.7%	78.2%	82.3%	86.8%
A_6	33.8%	37.6%	64.6%	63.6%	75.1%	74.1%	79.6%	87.3%
A_7	63.3%	63.1%	79.0%	81.8%	85.6%	87.2%	88.8%	90.4%
A_8	40.3%	41.3%	53.8%	56.7%	58.0%	63.0%	66.8%	81.2%
A_9	47.2%	51.2%	69.0%	67.8%	74.8%	74.0%	76.7%	87.1%

The worst *The second best* *The best*

Fig. 9. (A) Performance matrices of PUAM: A.1 Based on SE-DPS, 4 DPN and 5 CUN and A.2 Based on NE-DPS, 4 DPN and 5 CUN; (B) Performance matrix of EADM: B.1 Overall Performance, and B.2 F-Scores. Note that: ACC.: Accuracy, TT: Training Time

TABLE V
THE SUMMARY OF THE EVALUATION DATASET OF EADM

ID	Category	Description	Sample #
A_1	Mobility	Mobility related activity, e.g., taking a bus.	196
A_2	Study	Study related activities, e.g., reading a book.	267
A_3	Sport	Sport related activity, e.g., going to gym.	311
A_4	Service	Taking healthcare, personal care or other general services	408
A_5	Working	Work related activities, e.g., sending product promotions	431
A_6	Shopping	Shopping related activities, e.g., buying clothes	412
A_7	Entertaining	Entertaining related activities, e.g., going to a cinema	475
A_8	Socializing	Sharing personal feelings and meeting friends	572
A_9	Food	Food related activities, e.g., taking lunch	650
		Total # of training samples:	**3,722**

about the project testbed and key learnings of SSOA are discussed.

A. Evaluation of two analysis mechanisms

1) Evaluation data: A basic evaluation dataset is prepared with people-sensing App records, Instagram Feeds, and Tweets collected by ADC in December 2016. In total, it has around 1 million records, specifically 4k from App, 136k from Twitter and 810k from Instagram. The whole dataset will be used to test the performance of PUAM. Then a cleansed dataset with records generated in the project testbed area is used as the monthly document to define a category of activities and prepare a set of training samples as summarized in Table V, based on which, the performance of EADM is evaluated.

2) Performance of PUAM: The data processing performance of PUAM mainly relies on the scalable DBSCAN algorithm. Therefore, the evaluation is made by measuring influences of three key components, namely a) Data Partition Strategy (DPS), b) Data Partition Number (DPN), and c) Computing Unit Number (CUN). Accordingly, following configurations are made:

- DBSCAN Parameters : ϵ is 100m and $MinPts$ is 10.
- Data Partition Strategy (DPS): Two strategies are tested as Space-Even DPS (SE-DPS) and Number-Even DPS (NE-DPS) to split data with an even space coverage, and an even record number respectively;
- Data Partition Number (DPN): 4 DPNs are used, noted as DPN_X, $X \in \{3, 6, 9, 12\}$;
- Computing Unit Number (CUN): 5 CUNs are used, noted as CUN_Y, $Y \in \{1, 3, 6, 9, 12\}$, and all computing units have the same computation power.

Through a full combination of the above components, PUAM runs to process everyday data of the evaluation dataset. First, as shown in Figure 9 (A.1), SE-DPS gets its best performance in CUN_12 from 45.7s by using DPN_3 to 12.5s by using DPN_12. Even though the performance improves 76.7% comparing to the baseline, SE-DPS is inefficient in balancing the workload of computing units for optimal performance. In contrast, as shown in Figure 9 (A.2), NE-DPS can improve the performance dramatically with the increase of DPN and CUN. However, as shown by the red arrows, the increase of data partitions may also affect the performance due to the increased data processing and transmission time. In general, the optimal performance can be achieved when the total number of PU and AU is not less than a DPN, e.g., the best performance 4.7s is achieved in CUN_12 with DPN_12, which shows that PUAM is about 11 times faster than the conventional DBSCAN to process the same data.

3) Performance of EADM: Based on the evaluation dataset listed in Table V, EADM is compared with conventional methods such as J48, Naive Bayes, RF (Random Forest) and SMO (Sequential Minimal Optimization), and deep learning methods such as Sentence2Vec [24], CNN [25] and FastText.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2939496, IEEE Internet of Things Journal

A

1. Indicator: SNV

#	Place Name	SNV
1	Downtown	2,488
2	Orchard Rd	980
3	Vivo City (+Harbourfront MRT)	222
4	**Jurong East** (Transportation Hub + Shopping Malls)	**167**
5	Tampines (Transportation hub +Shopping Malls)	99
6	Novena area (MRT+Hospital +Shopping mall)	65
7	Boon Lay (MRT+Library Shopping mall)	47
8	Woodland (Transportation hub+ Shopping mall+Libary)	44
9	Buona Vista (MRT+Shopping mall)	43
10	Bishan (MRT+Library +Sport center)	30

2. Indicator: SP

#	Place Name	SP
1	Downtown	4.27‰
2	Orchard Rd	1.03‰
3	**Jurong East** (Transportation hub + Shopping malls)	**0.19‰**
4	Vivo City (+Harbourfront MRT)	0.18‰
5	Novena area (MRT+Hospital +Shopping mall)	0.13‰
6	Bishan (MRT+Library +Sport center)	0.08‰
	Buona Vista (MRT+Shopping mall)	0.08‰
8	Woodland (Transportation hub+ Shopping mall+Libary)	0.06‰
9	Tampines (Transportation hub + Shopping malls)	0.05‰
10	Clementi (MRT+Shopping mall)	0.04‰

3. Indicator: DP

#	Place Name	DP
1	Tampines (Transportation hub +Shopping Malls)	8.76
2	Vivo City (+Harbourfront MRT)	5.28
3	Choa Chu Kang (MRT+Shopping mall)	5.12
	Boon Lay (MRT+Library Shopping mall)	5.12
5	Orchard Rd	4.07
6	Sembawang (MRT+Library +Shopping mall)	3.90
7	**Jurong East** (Transportation Hub + Shopping Malls)	**3.85**
8	Ang Mo Kio (MRT+Shopping mall)	3.78
9	Woodland (Transportation hub+ Shopping mall+Libary)	3.05
10	Bedok (Transportation hub +Shopping mall)	2.99

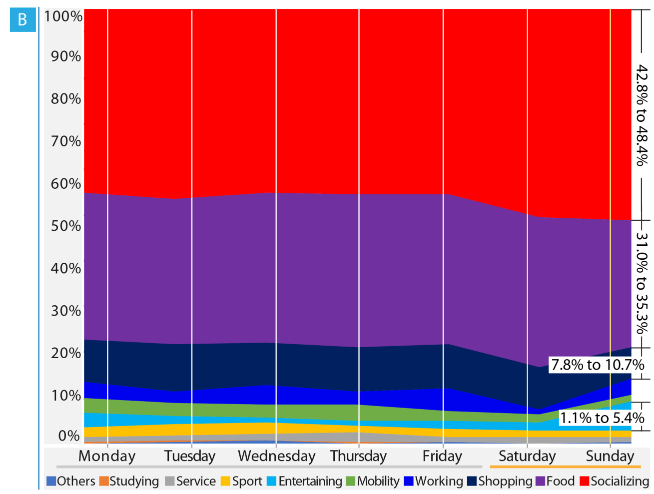


Fig. 10. The derived place design contexts: (A) Three top 10 lists of Singaporean regional centers and (B)The weekly activity distribution of the testbed

In order to avoid variations, a) all the algorithms run in a computer with 2GB memory and 2.5GHz 8 core CPU, b) 10 fold cross-validation is used to measure the performance, c) all conventional methods use N-gram ($0 < N < 4$) as classification features, d) all deep learning methods use a pre-trained local word model, and e) accuracy and training time are the indicators showing overall performance, and f-score is the indicator showing detailed performance per category,

As shown in Figure 9 (B.1), EADM achieves the highest accuracy of about 87.6%, which is about 9% higher than the second best method CNN. Then, conventional methods generally have a shorter training time (TT) than deep learning methods. However, comparing to the best two methods in TT, namely SMO, and Bayes, EADM can train a more accurate classifier in an acceptable time about one minute. Therefore, EADM outperforms rest methods in overall performance. Moreover, according to the f-score matrix shown in Figure 9 (B.2), EADM also overcomes other methods with the highest and more balanced f-scores in all 9 activity categories. Above all, it shows that EADM inherits the advantages of the two activity classifiers based through ensemble step, and overcomes compared methods with the highest classification accuracy and f-score per category.

B. Place Design Context

Based on the place utilization indicators as of the average number of visitors per day (ANV), the size of place (SP) against the land area of Singapore, and the density of people (DP), three top 10 lists of Singaporean regional centers are created as shown in Figure 10 (A). The project testbed Jurong East ranks 4th, 3rd and 7th in the ANV, SP, and DP respectively. The rankings show that it 1) serves as an important regional center together with these ones in downtown areas with a high visitor number; 2) hosts many facilities (e.g., shopping malls, MRT station, bus station, and hospital) with a large influence area, and 3) can further grow to host more facilities and serve more people comparing to other dense regional centers, such as Tampines and Vivo City.

As for the place utilization of the testbed Jurong East, a weekly activity pattern is analyzed as shown in Figure 10 (B). We can see that the top three activities are “Socializing”, “Food”, and “Shopping”, which remain high and stable proportions throughout a week. Moreover, an increase in “Socializing” and “Entertaining” during the weekend is observed, as more people visit the testbed to spend their spare time. This weekly pattern reveals that the testbed serves well as a regional center to support people’s daily lives not only as an attraction for shopping, but also as a venue for socializing and entertaining.

C. Key learnings of SSOA

After the discussion of SSOA and its first application IDP, some key characteristics of SSOA can be summarized as:

1) A common reference to utilize IoT and user systems in various SC domains: SSOA defines a three-tier architecture as a common reference to utilize various IoT and user systems for the implementation of smart services. So far, it has been applied for services in two SC domains, namely IDP for urban planning and Future Mobility Sensing [26] for urban mobility. The value of SSOA can be further explored by implementing more smart services in various SC domains to break the closeness of SC solutions and catalyze the “smart” transformation of cities.

2) A flexible mechanism to harness multi-source and multi-modal data: As demonstrated by IDP, SSOA provides a flexible mechanism to tackle issues in processing massive and heterogeneous data from both large scale IoT systems and user systems by using several dedicated components, i.e., the pair of “Adaptee” and “Adapter”, Adaptive Data Collector, Interconnected Data Model, Multi-dimension and Multi-measure Query, and Data Endpoints. By collaborating these components, multi-source and multi-modal data can be fused, mined and distributed for a comprehensive usage.

3) A unified framework to define and reuse system components: SSOA provides a framework with reusable IoT components and service modules to enable not only the solution-level redeployment, but also component-level reuse.

Such that, SSOA can reduce service implementation costs and improve service quality by reusing well developed and tested IoT components and modularized services, and adopting best practices that are seldom shared and exchanged before.

4) A scalable architecture to support synergetic orchestration: Comparing to conventional service/software architectures, SSOA breaks down vertical and horizontal orchestration barriers to enable a synergetic orchestration among the three elements. As demonstrated by the multi-source data processing flow, SSOA can scale up IoT services easily to incorporate more objects, harness diverse multi-source data, deploy concurrent micro-services and generate value proposition.

VI. Conclusion

As a solution to foster the development of SC and the usage of IoT by the synergetic orchestration among objects, data and services, this paper proposes SSOA with Cloud of Objects (COO), Cloud of Data (COD) and Cloud of Services (COS) to tear down orchestration barriers 1) vertically by a pair of "Adaptee" and "Adapter" between COO and COD, and a combination of data endpoints and management APIs between COD and COS; and 2) horizontally by the design of a common data access interface in COO, an interconnected data model in COD, and a pool of reusable micro-services in COS. Such that, SSOA enables a multi-source data processing flow to efficiently and effectively support the synergetic orchestration by incorporating multiple IoT and user-related objects, fusing heterogeneous data, and collaborating modularized services for value propositions in various SC domains.

As the first application of SSOA, this paper also presents the Informed Design Platform (IDP) with 1) the COO tier cooperating 5 IoT and user-related objects, 2) the COD tier integrating multi-source and multi-modal data according to a concrete interconnected data model with three common analysis dimensions (ADs) (i.e., place, time and people ADs) and several analysis measures (AMs) (e.g., place utilization, activity, etc.), and 3) the COS tier forming a pool of reusable service modules to create value propositions of the IDP Place Design Support Service. Moreover, as an instance of multi-source data processing flows in IDP, a place utilization and activity analysis flow is discussed to reveal the place design context of the project testbed Jurong East by using the data from IoT devices and user-related systems. It shows that Jurong East is a growing regional center with a high people visiting rate, a large influence area, and a moderate people density, and also an important regional center as an attraction for shopping and a venue for socializing and entertaining.

Additionally, in order to illustrate how machine learning methods can be utilized to support a comprehensive data fusion of massive and heterogeneous geo-referenced multi-source data, this paper proposes two dedicated analysis mechanisms, i.e., Place Utilization Analysis Mechanism (PUAM) and Ensemble-based Activity Detection Mechanism (EADM). As evaluated, PUAM running in a distributed environment can dramatically improve the performance of geospatial clustering about 11 times from the baseline 53.6s to 4.7s, and EADM measuring not only syntactic and semantic information of social messages but also functional information of enclosed POIs to detect place activities outperforms other compared methods with the highest accuracy about 87.7%, and also highest and more balanced f-scores in all activity categories.

In the future, first, SSOA will be used to support smart services in other domains, such as smart mobility [26]. Second, a standardized orchestration protocol will be studied to implement an automatic service orchestration engine that can create value propositions by assembling related IoT components and service modules. Third, more multi-source data processing flows will be implemented in IDP to reveal place design context in other aspects (e.g., public sentiment, accessibility, etc.), and accordingly more machine learning mechanisms will be studied to support the comprehensive data fusion. Finally, PUAM and EADM will be enhanced to analyze temporary places that may contain temporal but regular place usage needs, and long-term place utilization changes by measuring time series phenomena, such as concept drift.

Acknowledgment

The findings published in this paper are based on the project "Liveable Places: A Building Environment Modelling Approach for Dynamic Place Making", in principal collaboration with the Singapore Housing & Development Board (HDB, Building & Research Institute) and Urban Redevelopment Authority (URA), supported by a research grant (Grant No. SUL2013-5) from the Ministry of National Development Research Fund on Sustainable Urban Living.

References

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.

[2] N. Ianuale, E. Capobianco, and D. Schiavon, "Smart cities, big data and communities: Reasoning from the viewpoint of attractors," *IEEE Access: Smart Cities*, vol. 4, pp. 41–47, 2016.

[3] L. Wang, S. Hu, G. Betis, and R. Ranjan, "A computing perspective on smart city [guest editorial]," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1337–1338, 2016.

[4] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1125–1142, 2017.

[5] H. Yue, L. Guo, R. Li, H. Asaeda, and Y. Fang, "Dataclouds: Enabling community-based data-centric services over the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 5, pp. 472–482, 2014.

[6] B. Cheng, G. Solmaz, F. Cirillo, E. Kovacs, K. Terasawa, and A. Kitazawa, "Fogflow: Easy programming of iot services over cloud and edges for smart cities," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 696–707, 2018.

[7] M. Handte, S. Foell, S. Wagner, G. Kortuem, and P. J. Marrón, "An internet-of-things enabled connected navigation system for urban bus riders," *IEEE internet of things journal*, vol. 3, no. 5, pp. 735–744, 2016.

[8] F. Delmastro, V. Arnaboldi, and M. Conti, "People-centric computing and communications in smart cities," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 122–128, 2016.

[9] L. You, G. Motta, K. Liu, and T. Ma, "City feed: A pilot system of citizen-sourcing for city issue management," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 7, no. 4, p. 53, 2016.

[10] J. Venkatesh, B. Aksanli, C. S. Chan, A. S. Akyurek, and T. S. Rosing, "Modular and personalized smart health application design in a smart city environment," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 614–623, 2018.

[11] T. Hashimoto, D. Shepard, T. Kuboyama, and K. Shin, "Event detection from millions of tweets related to the great east japan earthquake using feature selection technique," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, 2015, pp. 7–12.

[12] M. Abu-Elkheir and H. S. Hassanein, "Orchestrating access to smart city services," in *Innovations in Information Technology (IIT), 2015 11th International Conference on*. IEEE, 2015, pp. 1–6.

[13] M. Stecca, C. Moiso, M. Fornasa, P. Baglietto, and M. Maresca, "A platform for smart object virtualization and composition." *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 604–613, 2015.

[14] M. Vögler, J. M. Schleicher, C. Inzinger, S. Dustdar, and R. Ranjan, "Migrating smart city applications to the cloud," *IEEE Cloud Computing*, vol. 3, no. 2, pp. 72–79, 2016.

[15] J. M. Schleicher, M. Vögler, S. Dustdar, and C. Inzinger, "Enabling a smart city application ecosystem: Requirements and architectural aspects," *IEEE Internet Computing*, vol. 20, no. 2, pp. 58–65, 2016.

[16] V. Gutiérrez, D. Amaxilatis, G. Mylonas, and L. Muñoz, "Empowering citizens toward the co-creation of sustainable cities," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 668–676, 2018.

[17] L. You, B. Tuner, and H. Xing, "Harnessing multi-source data about public sentiments and activities for informed design," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 343–356, Feb 2019.

[18] B. P. L. Lau, N. Wijerathne, B. K. K. Ng, and C. Yuen, "Sensor fusion for public space utilization monitoring in a smart city," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 473–481, 2018.

[19] F. Paganelli, S. Turchi, and D. Giuli, "A web of things framework for restful applications and its experimentation in a smart city," *IEEE Systems Journal*, vol. 10, no. 4, pp. 1412–1423, 2016.

[20] R. Liu, C. Yuen, J. Zhao, J. Guo, R. Mo, V. N. Pamadi, and X. Liu, "Selective ap-sequence based indoor localization without site survey," in *VTC2016-Spring, Nanjing, China*, 2016.

[21] I. Vilajosana, J. Llosa, B. Martinez, M. Domingo-Prieto, A. Angles, and X. Vilajosana, "Bootstrapping smart cities through a self-sustainable model based on big data flows," *Communications Magazine, IEEE*, vol. 51, no. 6, pp. 128–134, 2013.

[22] L. You and B. Tuncer, "Exploring the utilization of places through a scalable activities in places analysis mechanism," in *2016 IEEE International Conference on Big Data (Big Data)*, Dec 2016, pp. 3563–3572.

[23] E. Grave, T. Mikolov, A. Joulin, and P. Bojanowski, "Bag of tricks for efficient text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL*, 2017, pp. 3–7.

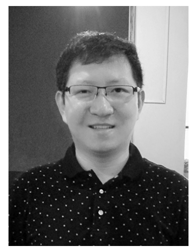
[24] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of The 31st International Conference on Machine Learning (ICML-14)*, vol. 14, 2014, pp. 1188–1196.

[25] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY: ACM Press, Aug. 2015, pp. 959–962.

[26] L. You, F. Zhao, L. Cheah, K. Jeong, C. Zegras, and M. Ben-Akiva, "Future mobility sensing: An intelligent mobility data collection and visualization platform," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov 2018, pp. 2653–2658.

Dr. Linlin You is a senior postdoctoral associate at the Singapore-MIT Alliance for Research and Technology and a research affiliate at the Intelligent Transportation System Lab at Massachusetts Institute of Technology. He was a research fellow at the Architecture and Sustainable Design Pillar of Singapore University of Technology and Design. He received his PhD in Computer Science from the University of Pavia, his dual masters with honor in Software Engineering and Computer Science from the Harbin Institute of Technology and the University of Pavia respectively, and his BSc in Software Engineering from the Harbin Institute of Technology. He published more than 30 journal and conference papers in research fields as Smart Cities, Service Orchestration, Multi-source and Multi-modal Data Fusion, and Social/Crowd Data Mining.

Dr. Bige Tunçer is an associate professor of design computation and the associate head of pillar at the Architecture and Sustainable Design Pillar of Singapore University of Technology and Design. She received her PhD from the Delft University of Technology, her MSc from the Carnegie Mellon University, and her BArch from the Middle East Technical University. She was an assistant professor at TU Delft, a visiting professor at the Chair of Information Architecture at ETH Zurich, and a visiting scholar at MIT.



Dr. Rui Zhu is a postdoctoral associate in the Senseable City Laboratory of Future Urban Mobility, Singapore-MIT Alliance for Research and Technology. His research interests are in geospatial data processing, urban mobility, ride sharing, and urban heat islands. Dr. Zhu holds a Ph.D. in Geo-Informatics from the Hong Kong Polytechnic University (Hong Kong SAR), with an exchange study at the Universite Laval (Canada). Before that, Dr. Zhu obtained an MSc in Geodesy and Geo-Informatics from the KTH - Royal Institute of Technology (Sweden), and a BSc in Geographical Information Science from the Nanjing Normal University (China).

Hexu Xing is a research assistant at the Architecture and Sustainable Design Pillar of Singapore University of Technology and Design. He received his dual master's degree in Computer Engineering from the University of Pavia, Italy, and Software Engineering from the Northeastern University, China. His research interest is about social network data mining.

Dr. Chau Yuen received the BEng and PhD degree from Nanyang Technological University (NTU), Singapore, in 2000 and 2004 respectively. Dr. Yuen was a Post Doc Fellow in Lucent Technologies Bell Labs, Murray Hill during 2005. During the period of 2006 2010, he worked at the Institute for Info-comm Research (I2R, Singapore) as a Senior Research Engineer, where he was involved in an industrial project on developing an 802.11n Wireless LAN system, and participated actively in Long Term Evolution (LTE) and LTE-Advanced (LTEA) standardization. He joined the Singapore University of Technology and Design from June 2010, and received IEEE Asia-Pacific Outstanding Young Researcher Award in 2012. Dr. Yuen serves as an Editor for IEEE Transactions on Communications and IEEE Transactions on Vehicular Technology. He has 2 US patents and published over 300 research papers at international journals or conferences.