A Smart City Data Fusion Framework Based on Multi-Source Data: Postprint

Authors: Ma Jie, Ge Yan, Pu Hongyu, Zhang Yunkai

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Effectively realizing multi-source heterogeneous data fusion in smart cities constitutes a core challenge that must be addressed. [Method/Process] Based on defining the urban data resource system, this study proposes a semantic-oriented metadata model and, combined with user demand classification, constructs a smart city data fusion framework based on multi-source data. [Results/Conclusion] Relying on the smart city data operation center, this paper proposes an implementation path for user data service recommendation oriented toward smart cities; finally, it puts forward development suggestions for smart city data fusion: improving smart city multi-source data fusion standards; actively formulating data openness policies; and exploring solutions for personal data collection and storage protection.

Full Text

A Smart City Data Fusion Framework Based on Multi-Source Data

Ma Jie[1,2], Ge Yan[1,3], Pu Hongyu[1], Zhang Yunkai[1]

[1]School of Management, Jilin University, Changchun 130022
[2]Center for Information Resources Studies, Jilin University, Changchun 130022
[3]School of Computer Science and Technology, Beihua University, Jilin 132000

Abstract

[**Purpose/Significance**] Effectively achieving the fusion of multi-source heterogeneous data represents a core challenge that smart cities must address. [**Method/Process**] Building upon a delineation of the urban data resource system, this study proposes a semantic-oriented metadata model and constructs a smart city data fusion framework based on multi-source data, integrated with user demand classification. [**Result/Conclusion**] Relying on the smart city data operation center, the paper proposes an implementation path for user data service recommendation in smart cities. Finally, it offers development recommendations for smart city data fusion: improving multi-source data fusion standards for smart cities; actively formulating data openness policies; and exploring solutions for personal data collection and storage protection.

Keywords: Smart City, Data Fusion, Metadata, User Requirements

---

1. Introduction

Smart cities represent an advanced form of urban development built upon urban informatization [1]. The "Guiding Opinions on Promoting the Healthy Development of Smart Cities," jointly issued by eight ministries and commissions, proposes that "by 2020, a batch of smart cities with distinctive characteristics will be established" [2]. The "13th Five-Year Plan for National Economic and Social Development" explicitly calls for strengthening modern information infrastructure and building smart cities. The "New Smart City Development Report 2017" indicates that China's new smart city construction is developing in clusters. At the National Cybersecurity and Informatization Work Conference in April 2018, President Xi Jinping emphasized the need to promote in-depth development of digitalization, big data, and artificial intelligence, with big data applications providing powerful technical support for smart city development.

Currently, the industry has not yet reached a unified definition of the smart city concept. Wang Jingyuan defines a smart city as "a new urban form based on new-generation information technology that monitors, analyzes, and utilizes data from various parts of the city to achieve thorough perception of the urban living environment, comprehensive regulation of urban resources, convenient operation of all urban aspects, and harmonious win-win relationships between people and the city" [3]. Li Chongzhao believes that the connotation of smart cities includes the foundational role of information and communication technology, economic development and innovation encouragement, and promotion of resource sharing and collaborative operations among various departments [4]. Although domestic scholars interpret smart cities from different perspectives, their core concepts are basically consistent: using new-generation information technology to integrate urban operation core data and provide intelligent services to the public.

Smart cities represent an emerging interdisciplinary field. In traditional urban science, most scholars use urban operation data to identify urban functional areas, analyze urban transportation networks, and model urban resident behavior. In computer science, researchers have proposed smart city technical system frameworks centered on data mining, processing, and analysis technologies. In management science, scholars focus on the vision and strategic goals of smart cities [5] and analysis of factors influencing future urban development [6]. In industry, the emphasis is on specific project development and application in public services, business, and energy sectors [7].

Urban data encompasses spatial and temporal dimensions, and different data may require entirely different processing methods and standards due to differences in form and structure. For example, traffic condition data may include static road monitoring information, Weibo traffic reports, and traffic accident records. Although these three data forms are different, they all address the same traffic flow issue. Furthermore, through investigating domestic smart city construction, we find that most projects build upon existing urban informatization achievements, facing certain technical barriers and sharing permission issues in data sharing and integration. Wang Guangbin, through a literature review of foreign smart city research, concludes that international studies pay more attention to technological progress and data integration and fusion issues [8]. In 2018, the National Smart City Standardization Group released the "Smart City Data Fusion" standard document, which specifies conceptual models and data encoding specifications for data collection, organization, interconnection, and services. This study, based on clarifying the urban data resource system, proposes a semantic-oriented metadata model and constructs a smart city data fusion framework based on multi-source data and user demand classification.

---

2. Urban Data Resource System

The urban data resource system is a complex system that integrates various domain data within urban spatiotemporal scope based on their intrinsic relationships, using geospatial data as the unified carrier. Government, enterprise, and public entities constitute the main bodies of smart cities. From a systematic perspective, using geospatial data as the foundation, we construct a smart city data resource system based on urban entities, as shown in Figure 1 [Figure 1: see original paper].

The urban data resource system in Figure 1 includes: (1) **Basic Geographic Information Element Data**: Data related to geographic location that serves as the spatial data basis for standardized information processing [18], including imagery, vector maps, and topographic maps. Building upon the national standard "GB/T 13923-2006 Classification and Codes for Basic Geographic Information Elements," some cities have developed local standards tailored to their urban planning and surveying characteristics to enrich urban basic geographic information content. (2) **Government Data**: The collective term for data resources generated inside or outside government that affect government activities, public affairs, public life, and urban operation [19], including laws and regulations, policy documents, departmental reports, public registration information, environmental and meteorological information, research databases, and statistical compilations from emergency management, transportation, public security, environmental protection, meteorology, water affairs, and other government departments. (3) **Enterprise Data**: Data related to business operations, including data generated during product R&D, design, manufacturing, marketing, and capital flow. (4) **Public Data**: Data generated by individuals in daily life, including mobile phone data, LBS location service data, and passenger flow data. These social activity data depict user activity patterns. Mobile phone data includes call records, GPS positioning information, and internet browsing records; LBS location service data can obtain location information of mobile terminal users and serves as a supplement to POI data; passenger flow data contains information about urban population movements.

---

3. Semantic-Oriented Smart City Metadata Model

Metadata is structured data that describes the attributes of a particular type of resource. A standardized smart city metadata model forms the basis for interoperability among various urban application systems, can resolve heterogeneous data conversion, enable resource aggregation on the same theme, and provide data services for urban decision-makers. Metadata service providers, primarily data producers and owners, publish metadata services to the city's UDDI registration center [21]. This study defines the metadata model as a six-tuple MD = {S, E, A, I, R, C} [22].

(1) Data Source S (Source): Data from education, transportation, health, taxation, housing and construction, public security, civil affairs, meteorology, water affairs, and other departments. The data source set is represented as S = {S_1, S_2, ..., S }, where S (1 < i < n) represents the i-th data source.

(2) Entity Type Set E (EntityType): The general term for entity sets with the same attributes. Entities include people, objects, and spatiotemporal entities. People (Agent) is a general term for government, enterprise, and public entities—data holders capable of autonomous activity. Objects include natural geographic entities (such as mountains, rivers, lakes) and artificially constructed geographic entities (such as buildings, roads, bridges, and streets). Spatiotemporal entities (TemporalSpatialEntity) refer to objects with spatiotemporal multidimensional characteristics; for example, various sensors (temperature, humidity, etc.) within urban scope can provide citizens' travel information, air quality monitoring information, and weather conditions. As shown in Figure 2 [Figure 2: see original paper].

(3) Entity Attributes A: The entity attribute set A = {a_{11}, a_{12}, ..., a }, where a (1 < i < m, 1 < j < k) represents the j-th attribute of the i-th entity. Taking sensing devices as an example: sensing devices are physical devices that can perceive changes in external information in real time and transmit acquired information to other devices, such as sensors, GPS positioning devices, video monitoring devices, and RFID devices. The entity attributes are shown in Table 1 .

(4) Instance Set I: A collection of entity class objects. An instance is an identifiable object that exists in the real world, represented as I = {I_1, I_2, …, I }, where I (1 < i < m) represents the i-th entity.

(5) Binary Semantic Relations R Between Instances: Based on the PAS182 conceptual model for smart city data interoperability, we summarize semantic relations between data entity objects as shown in Table 2 , including hierarchical relations, conceptual relations, physical relations, spatial relations, functional relations, and business relations. Relations between entity class objects can all be described through basic data and semantic relations. For example, "contain" describes inclusion relations between instances—the Changchun Municipal Government includes Chaoyang District Government, Nanguan District Government, Erdao District Government, Kuancheng District Government, and Lvyuan District Government.

(6) Context C (Context): This primarily expresses the weather environment, scene situation, and background information where entities are located. Contextual relations can effectively resolve data conflicts [24]. Entity attributes obtained from different data sources, after contextual confirmation, can represent the same entity, thereby improving data fusion quality. As shown in Figure 3 [Figure 3: see original paper], in contexts C_1, C_2, C_3, C_4, and C_5, entity attributes a_{11} and a_{12} are extracted from data sources S_1, S_2, and S_3, and then determined to belong to the same entity e_1.

---

4. Construction of Smart City Data Fusion Framework Based on Multi-Source Data

4.1 Smart City User Demand Classification 4.1.1 Kano User Demand Classification Model: The Kano model categorizes demands into five types: basic needs, performance needs, excitement needs, indifferent needs, and reverse needs. (1) **Basic Needs**: Fundamental user requirements for a product. When these needs are met or exceeded, users may not necessarily feel satisfied, but when unmet, users will definitely be dissatisfied. (2) **Performance Needs**: Needs where user satisfaction is proportional to the degree of fulfillment. Meeting these needs significantly increases user satisfaction. (3) **Excitement Needs**: Needs that users do not explicitly expect. When met, they generate high satisfaction; when absent, they do not cause dissatisfaction. (4) **Indifferent Needs**: Needs that, whether provided or not, have no impact on user experience or satisfaction. (5) **Reverse Needs**: Quality characteristics that cause strong dissatisfaction or low satisfaction. These needs lead to decreased user satisfaction. As shown in Figure 4 [Figure 4: see original paper].

4.1.2 Smart City User Demand Classification Based on Kano: Government is the operation manager of smart cities, while the public is the target of smart city management and services. User demand constitutes the driving force of urban development [25], and user needs and feedback cannot be ignored—

meeting user demand is key to smart city construction. However, user needs are dynamic, multifaceted, and uncertain, and smart cities encompass numerous specific applications across many domains, making precise description of specific needs difficult. Using the Kano model and combining it with five stages of individual life cycles (infancy, childhood, adolescence, adulthood, and old age), we gradually refine to obtain demand groups with different functional attributes (see Figure 5 [Figure 5: see original paper]), developing specific products covering different industry domains, including smart education products, smart healthcare products, smart transportation products, and smart community products.

4.2 Construction of Multi-Source Data Fusion Framework The service subjects of smart city data are government, enterprise, and public entities. The main objectives of the smart city data fusion framework are: at the micro level, through collection, processing, and analysis of basic geographic information data, government data, enterprise data, and individual data, to achieve public opinion monitoring, early warning monitoring, and problem localization; at the meso level, to establish an urban data fusion platform that proactively pushes smart city-related data services to achieve people-benefiting services and precise urban governance; at the macro level, to provide important significance for the formulation of various urban policies. Therefore, adopting a distributed storage structure based on big data technology, we propose constructing an integrated data fusion framework based on the public information service platform, government intranet information exchange platform, enterprise information service platform, and spatial positioning information platform, as shown in Figure 6 [Figure 6: see original paper].

Combined with user demand classification and following the enterprise-level big data technology framework, the specific implementation of the smart city data fusion framework is divided into four steps:

(1) Data Collection: Main data sources include basic geographic information data, government data, enterprise data, and public data. The data collection process includes offline collection and real-time collection. Offline collection uses SQL collection adaptation and Sqoop collection. SQL collection adaptation primarily collects database data through SQL. Sqoop, as a tool for transferring data between Hadoop and relational databases, imports data from relational database management systems into the Hadoop Distributed File System (HDFS) and completes data transformation tasks under the Hadoop MapReduce framework, providing basic data sources for data fusion. Real-time data collection uses script collection and Flume. Script collection mainly uses Shell scripts or APIs to obtain data from databases or websites, converting unstructured and semi-structured data into structured data and storing them as unified local file data. Flume, as a distributed real-time log collection system, captures massive log information from different servers and pushes it to the Kafka distributed message management system.

(2) Data Storage: Includes offline data storage and real-time data storage. Offline data storage uses HDFS or HBase to store offline collected data. Real-time data storage uses Redis or KafkaServer.Redis distributed NoSQL databases. Redis serves as a cache database for users to store streaming data; KafkaServer mainly stores data collected by Kafka. For non-real-time structured data, such as urban geographic information data with fixed table structures generally queried using SQL, storage is in the Hive database. For real-time data, such as real-time urban road vehicle operation data collected at second-level rates and event log data with strong suddenness, which require high data throughput performance and have relatively simple access patterns (generally by time series and object ID using key-value pairs), storage is in the HBase database. For semi-structured data such as video and images, files are stored in the HDFS distributed file system.

(3) Data Scheduling: The data scheduling layer mainly includes Oozie task scheduling and Zookeeper. Oozie is a workflow scheduling tool for managing and organizing Hadoop jobs. Zookeeper is a coordination tool for distributed applications that stores important information from various components and manages coordination between Kafka and Storm. Hue is a visualization tool in the Hadoop ecosystem that enables viewing the operational status of various big data components.

(4) Data Sharing: The application data center shares processed data with the public information service platform and government intranet information exchange platform. YARN, as a general resource management system, provides unified resource management and scheduling services for data applications and can improve cluster server utilization. The query data center is mainly composed of HBase, Phoenix, and YARN, and the combination of these components can improve big data query speed.

Through these four steps, the smart city data fusion framework achieves the following functions: (1) By fusing various government data, it establishes big data decision analysis models for time series analysis and trend prediction of massive data, which can enhance urban management and planning capabilities, emergency event early warning capabilities, and provide data-oriented guidance for government policy formulation. (2) By integrating different sources and types of enterprise internal and external data, it helps understand market dynamics and technological frontiers, providing data support for scientific enterprise operation and decision-making. (3) It provides data recommendation services to the public in education, transportation, healthcare, tourism, and other areas. (4) It enables urban data visualization analysis. By integrating 2D maps, 3D real-world scenes, video data, government data, enterprise data, and public data within the same spatiotemporal system, it establishes a spatiotemporal multi-dimensional urban management system, thereby enabling real-time, intuitive visual urban management and improving refined urban governance capabilities.

4.3 Implementation Path for User Data Service Recommendation in Smart Cities Based on the multi-source data fusion framework, we establish a smart city data operation center responsible for collecting, managing, and sharing big data in urban life, establishing a paradigm for multi-departmental cooperation within the city that can improve the current insufficient horizontal collaboration among government departments. Grounded in smart city user demand classification and combined with demographic attributes and usage behavior preferences, we construct smart city user demand profiles to provide specific data recommendation services. Figure 7 [Figure 7: see original paper] illustrates three implementation paths for user data recommendation services: (1) Path 1 conducts coarse-grained matching between user demand information and smart city-related knowledge bases to provide smart city-related services to users; (2) Path 2 conducts fine-grained matching between user demand semantic information and knowledge units extracted from smart city knowledge bases, performing multidimensional data fusion on knowledge units to provide content-oriented data recommendation services to users; (3) Path 3 studies how to open data interfaces to resource owners and service providers, and provide corresponding data services based on user demand profiles.

---

5. Development Recommendations

The design of the smart city data fusion framework serves as top-level design for technical research and is crucial for building smart city big data management and analysis platforms. Additionally, it requires active cooperation from relevant departments to better provide data recommendation services to the public.

5.1 Improve Multi-Source Data Fusion Standards for Smart Cities and Accelerate Coordinated Development of Urbanization, Datafication, and Standardization

Smart city construction requires coordinated development of urbanization, datafication, and standardization. The "National New Urbanization Plan (2014-2020)" proposes accelerating smart city construction. In the big data environment, IoT, cloud computing, and sensor technology transform information circulating within urban scope into data, endowing it with data attributes and thus presenting a development trend of urban datafication. The "Smart City Data Fusion" standard document includes data encoding specifications, data collection specifications, and municipal infrastructure data element specifications. Based on this standard, we should further improve multi-source data fusion standards for smart cities, including entity data standardization, entity attribute standardization, and application context standardization, establishing data conversion standards and storage specifications to achieve cross-domain, cross-departmental, and cross-level data fusion. See Figure 8 [Figure 8: see original paper].

5.2 Actively Formulate Data Openness Policies and Establish Multi-Party Cooperation Mechanisms for Data Sharing Platforms

Government should formulate data openness policies that clearly define the scope of openable data. As data providers, they should strive to provide raw data sharing services while complying with policy requirements to ensure the completeness of data fields. As data recipients, they still bear the responsibility for ensuring information security under the premise of reasonable data use. Government provides decision-making optimization support, universities provide intellectual support, and enterprises provide technology and product support to establish a data sharing mechanism with stakeholder cooperation relationships.

5.3 Explore Personal Data Collection and Storage Protection Solutions and Encourage Citizen Participation in Smart City Construction

Personal data reflects real-life conditions and encompasses sensitive information such as clothing, food, housing, and transportation. In recent years, frequent personal privacy leakage incidents have caused panic about personal data privacy, and the resulting "chilling effect" may make the public reluctant to share personal data. Therefore, researching smart city personal data collection and storage protection solutions is imperative. Smart city construction should widely solicit citizen opinions. For example, Amsterdam outsources smart city application function design to citizens, who complete design proposals and even preliminary application development in community or group formats. China can learn from this model to encourage citizens to directly participate in smart application development, providing smart city services that meet market demands.

References

[1] Zhang Yuanhao, Zeng Zhenxiang. Literature review on urban informatization: from information port, digital city to smart city[J]. Information Science, 2015, 33(6): 131-137.

[2] Central People's Government of the People's Republic of China. Notice on Issuing the Guiding Opinions on Promoting the Healthy Development of Smart Cities by the National Development and Reform Commission, Ministry of Industry and Information Technology, Ministry of Science and Technology, Ministry of Public Security, Ministry of Finance, Ministry of Land and Resources, Ministry of Housing and Urban-Rural Development, and Ministry of Transport[EB/OL]. [2017-12-08]. http://www.gov.cn/gongbao/content/2015/content_{2806019}.htm.

[3] Wang Jingyuan, Li Chao, Xiong Zhang, et al. Data-centric smart city research review[J]. Journal of Computer Research and Development, 2014, 51(2):

239-259.

[4] Li Chongzhao, Liu Shuhua. Smart city: a new trend in China's urban governance[J]. E-Government, 2011(6): 13-18.

[5] Xu Qingrui, Wu Zhiyan, Chen Litian. Vision and architecture of smart cities[J]. Journal of Industrial Engineering and Engineering Management, 2012, 26(4): 1-7.

[6] Wu Xibo, Yang Zaigao. Smart city concept and future urban development[J]. Urban Development Studies, 2010, 17(11): 56-60.

[7] EU-Project SmartEnCity: Towards Smart Zero CO2 Cities across Europe[EB/OL]. [2017-12-08]. http://smartencitynetwork.eu/.

[8] Wang Guangbin, Cui Qinghong, Liu Huan. Current status analysis and enlightenment of foreign smart city research[J]. Urban Development Studies, 2015, 22(6): 40-46.

[9] Gaba V. Understanding smart cities: an integrative framework[C]//Proceedings of the 45th Hawaii international conference on system sciences. Piscataway: IEEE, 2012: 2289-2297.

[10] Nam T, Pardo T. Conceptualizing smart city with dimensions of technology, people, and institutions[C]//Proceeding of the 12th international digital government research conference: digital government innovation in challenging times. New York: ACM, 2011: 282-291.

[11] Huo Guoqing, Meng Guangjun, Wang Jinxiao, et al. Sublimation of information resource management thought[J]. Library and Information Service, 2002(4): 26-39.

[12] Chen Rui, Jia Xiaofeng, Zhao Yu. Information collaboration standard system for smart city operation management[J]. Journal of Urban Development Studies, 2015, 22(6): 40-46.

[13] Zheng Yu. Overview of urban computing[J]. Geomatics and Information Science of Wuhan University, 2015, 40(1): 1-13.

[14] Pan G, Qi G, Zhang W, et al. Trace analysis and mining for smart cities: issues, methods and applications[J]. IEEE communications magazine, 2013, 51(6): 120-126.

[15] Cao Yujie, Li Gang, Mao Jin, et al. Emergency information fusion oriented to the decision-making full process in big data environment[J]. Library and Information Knowledge, 2018(5): 95-104.

[16] Zhang Yi, Chen Yujun, Du Bowen, et al. Multi-mode data fusion model for smart cities[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(12): 2683-2690.

[17] Nesi P, Badii C, Bellini P, et al. Km4City smart city API: an integrated support for mobility services[C]//IEEE. 2016 IEEE international conference on

smart computing. USA: IEEE, 2016: 1-8.

[18] Yang Lina. Construction of spatiotemporal data warehouse for smart city data management and multidimensional decision-making[J]. Science of Surveying and Mapping, 2014, 39(8): 44-48.

[19] Yang Ruixian, Mao Chunlei, Zuo Ze. Comparative study on the status of government data openness at home and abroad[J]. Journal of Intelligence, 2016, 35(5): 167-172.

[20] Zhang Han, Wang Zhong. Comparative study on foreign government open data[J]. Journal of Intelligence, 2015(8): 142-146.

[21] Wang Hu, Li Qi, Dong Baoqing, et al. Constructing metadata service system for digital cities[J]. Computer Science, 2003, 30(8): 85-87.

[22] Huang Hongbin, Zhang Weiming, Deng Su, et al. Research and implementation of metadata model for semantic information sharing[J]. Computer Science, 2008(4): 124-128.

[23] Jia Xiaofeng, Liang Zhengli, Ren Jinluan. Multi-source information collaboration: application and evolution of city and regional level big data[M]. Beijing: Tsinghua University Press, 2016.

[24] Zhang Yongxin, Li Qingzhong, Peng Zhaohui. Two-stage data conflict resolution method based on Markov logic network[J]. Chinese Journal of Computers, 2012, 35(1): 101-111.

[25] Wang Jinzhu, Li Jiawei. Philosophical examination of smart cities[J]. Studies in Dialectics of Nature, 2018(11): 119-123.

Author Contributions

Ma Jie: Conceived the article framework and revised the manuscript.
Ge Yan: Wrote and revised the main content of the article.
Pu Hongyu: Wrote and revised the main content of the article.
Zhang Yunkai: Collected literature and translated English content.

Intelligent City Data Fusion Framework Based on Multi-Source Data

Ma Jie[1],[2], Ge Yan[1],[3], Pu Hongyu[1], Zhang Yunkai[1]

[1]School of Management, Jilin University, Changchun 130022
[2]Center for Information Resources Studies, Jilin University, Changchun 130022
[3]School of Computer Science and Technology, Beihua University, Jilin 132000

Abstract: [Purpose/significance] How to effectively realize data fusion of multi-source heterogeneous data in smart city is the core problem that must be solved in smart city. [Method/process] Based on the definition of urban data resource system, this paper proposes a semantic-oriented metadata model, and constructs a smart city data fusion framework based on multi-source data according to user needs classification. [Result/conclusion] Based on the smart city data operation center, this paper proposes the implementation path of user data service recommendation for smart city. Finally, it puts forward some suggestions for the development of smart city data fusion: improving the multi-source data fusion standard of smart city; actively formulating data opening policy; exploring personal data collection and storage protection scheme.

Keywords: intelligent city; data fusion; metadata; user requirement

Note: Figure translations are in progress. See original paper for figures.