Engineering 2 (2016) 171–178

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
iCity & Big Data—Review

Urban Big Data and the Development of City Intelligence

Yunhe Pan[a,*], Yun Tian[a,b,c], Xiaolong Liu[a], Dedao Gu[d], Gang Hua[d]

[a] Chinese Academy of Engineering, Beijing 100088, China
[b] College of Information Science and Technology, Beijing Normal University, Beijing 100875, China
[c] Department of Automation, Tsinghua University, Beijing 100084, China
[d] Ningbo Academy of Smart City Development, Ningbo, Zhejiang 315048, China

ARTICLE INFO

Article history:
Received 13 April 2016
Revised 16 May 2016
Accepted 26 May 2016
Available online 22 June 2016

Keywords:
Urban big data
City intelligence
Ternary space
Construction emphases

ABSTRACT

This study provides a definition for urban big data while exploring its features and applications of China's city intelligence. The differences between city intelligence in China and the "smart city" concept in other countries are compared to highlight and contrast the unique definition and model for China's city intelligence in this paper. Furthermore, this paper examines the role of urban big data in city intelligence by showing that it not only serves as the cornerstone of this trend as it also plays a core role in the diffusion of city intelligence technology and serves as an inexhaustible resource for the sustained development of city intelligence. This study also points out the challenges of shaping and developing of China's urban big data. Considering the supporting and core role that urban big data plays in city intelligence, the study then expounds on the key points of urban big data, including infrastructure support, urban governance, public services, and economic and industrial development. Finally, this study points out that the utility of city intelligence as an ideal policy tool for advancing the goals of China's urban development. In conclusion, it is imperative that China make full use of its unique advantages—including using the nation's current state of development and resources, geographical advantages, and good human relations—in subjective and objective conditions to promote the development of city intelligence through the proper application of urban big data.

1. Introduction

Amid China's rapid industrialization and urbanization, the rise in the population, manufacturing, and traffic of its cities is becoming increasingly intense and complex leading to a variety of urban diseases such as rapid population growth, traffic jams, environmental deterioration, housing shortages, employment problems, and public safety challenges. This is just a short list of the side effects of urbanization while there is a host of other less prominent policy problems facing Chinese policymakers. All of these factors have become serious constraints upon the healthy and sustainable development of China's urban ecosystems [1]. On one hand, the urban ecosystem is composed of urban infrastructure and diverse social environment among urban residents that is becoming increasingly intricate and scaling upon a daily basis; on the other, decision makers and administrators are not fully conscious of this complexity and are consequently deficient in the efficient management of this ecosystem. Modern cities have been upgraded to ternary spaces from dual spaces. The first-dimensional space is a physical space made up solely of a physical environment with all its resources in a natural state. The second-dimensional space contains a human society space shaped and sustained by the culture, norms, and social interactions of urban residents. A third-dimensional space, unlike the previous two, is a cyber space, which is comprised of computers, internet access, and the data flowing through these systems to informationized domain [2]. This new structural concept of urban life calls for new philosophies, theories, and practices for ana-

* Corresponding author.
E-mail address: zhs@cae.cn

http://dx.doi.org/10.1016/J.ENG.2016.02.003

lyzing the structures, economic development, and governance problems of urban life in a meaningful and systematical way, to better understand the new direction of the new urban landscape in context of this technological revolution.

The advent of "intelligent city" and "big data" provides new potential platforms for resolving various "urban diseases." Intelligent cities provide an artificial nervous system based upon the blueprints of traditional urban ecosystems, while urban big data describes the physical domain of real objects (buildings, cars, roads, and so on) and social domain of urban residents in real space, all of which has a virtual form that reflects the real forms in first-dimensional space and second-dimensional space. These virtual forms constitute the third-dimensional space: urban big data. Scholar Guojie Li, member of the Chinese Academy of Engineering (CAE), believes that the role of big data is similar to the "Honeybee Model," which is based on the value of having honey bees: The bees improve agricultural output of farms through the pollination of crop plants rather than the honey they produce [3]. Just like the human and non-human resources of cities, urban big data has become an important strategic resource for the development of intelligent city and strategic direction. As a city evolves toward informatization and intelligence, numerous information bases and data centers have been emerging, which should be properly interconnected to form urban big data. Urban big data can be converged, analyzed, and mined with depth via the Internet of Things, cloud computing, and artificial intelligence technology to help people understand the forces guiding the development for each layer of the urban system to assist the government and society with decision-making and urban planning to achieve the goal of intelligent administration of the city. Meanwhile, urban big data will bring about profound changes to the operating mode of various urban sectors, and speed the transformation and upgrading of traditional industries as well as the development of emerging industries. Thus, urban big data is speeding up the development of city intelligence. Specifically, city intelligence is implemented through the combined development of the Internet, Internet of Things, telecommunication networks, radio and television networks, and wireless broadband networks; it is characterized mainly by thorough integration of information technologies and comprehensive applications of big data; it focuses on intelligent technologies, intelligent industries, intelligent services, intelligent administration, and intelligent life; lastly, this transformation is committed to building a new form of urban development that is capable of self-correction and solving critical social, economic, and ecological problems in a more automated and timely fashion [2].

2. Overview of urban big data

2.1. Definition and features of urban big data

Urban big data is a massive amount of dynamic and static data generated from the subjects and objects including various urban facilities, organizations, and individuals, which have been being collected and collated by city governments, public institutions, enterprises, and individuals using a new generation information technologies. Big data can be shared, integrated, analyzed, and mined to give people a deeper understanding of the status of urban operations and help them make more informed decisions on urban administration with a more scientific approach, thereby optimizing the allocation of urban resources, reducing the operating costs of the urban system, and promoting the safe, efficient, green, harmonious, and intelligent development of the cities as a whole. In addition to their general features (e.g., volume, velocity, variety, veracity, and value), urban big data also has the additional features below:

(1) Hierarchy: For example, electronic medical records are categorized by hospital or region, while medical images are categorized in terms of individual medical devices and hospitals. Meanwhile, health data can be categorized in terms of individuals, hospitalized patients, communities, or health and anti-epidemic authorities. The hierarchy of urban big data deeply reflects the organizational hierarchy of a city's physical and social systems.

(2) Integrity: As an urban system evolves, the data coverage of each subsystem becomes increasingly broad. In recent years, for example, the inclusion of environmental protection data in China's cities has improved rapidly. Due to the rapid improvement of data integrity, urban big data has acquired the capacity to uncover the overall dynamics of urban development increasingly accurately.

(3) Correlation: Types of urban data are highly correlated with each other. For example, information about urban logistics is included not only in the data of logistics enterprises but also in the data of the manufacturing, commercial, and transport industries, and even in the financial industry. Such correlations can be used, not only for mutual corroboration, but also for cooperative reasoning and mining rules of cities operation.

Due to these general and unique features, urban big data must be applied using a new data processing technology—targeted data extraction based on a target-driven method [4]. The entire data process, from data acquisition to data processing to data modeling, is automated [5,6]. This occurs as follows: The first step is acquiring and storing the original data, including pattern extraction and filtering the data obtained from the desired data source according to the target requirements and then cleaning and preprocessing the acquired data (i.e., data filling, data optimization, data merging, data normalization, data consistency check, and preliminary organization of diverse data attributes), and establishing the dataset to be processed. The second step is processing and analyzing the dataset (including linear analysis, nonlinear analysis, factor analysis, sequential analysis, linear regression, variable curve analysis, and bivariate statistics), and then categorizing the data and analyzing the inter-data and inter-category relationships via the support vector machine (SVM), Naïve Bayes, random forest, and logistic regression. The third step is identifying the inherent relationships among the categorized data and uncovering the further patterns, rules, and knowledge via an artificial neural network, genetic algorithm, and cross-media algorithm. Finally, the relationships among the variables are explained in an interactive and visual way to express a deeper understanding of the results.

2.2. Categorization of urban big data

Urban big data describes the real-time status of various urban elements, including buildings, streets, pipelines, environments, enterprises, finance, commerce, products, markets, logistics, medicine, culture, education, traffic, public order, and population. As proposed in Ref. [7], urban big data can be categorized into five types: sensor data on urban infrastructure and moving objects, user data on society and humans, governmental administration data, customer and transaction record data, and arts and humanities data. Table 1 lists examples and user groups for the five types.

Urban big data can be categorized in more than one way. In essence, data categorization is not effective without a tree structure, but urban data information is organized using a mesh structure. Therefore, urban big data should be categorized according to the data processing method used and the application objective. The big data on China's cities is typically categorized using the three methods below.