SpoVis: Decision Support System for Site Selection of Sports Facilities in Digital Twinning Cities

Ke Zhang, *Member, IEEE,* Hao Chen, Hong-Ning Dai, *Senior Member, IEEE,* Hongbo Liu, Zhongrui Lin

***Abstract*—The site selection of sports facilities is a pivotal link in the construction of city livable environment and the development of sports business in digital-twinning cities. Recent years have witnessed data mining and visualization technologies bringing the convenience as well as opportunities for intelligent site selection. However, the lack of effective and reliable systematic analysis leads to difficulties in developing sports facilities planning schemes and constructing the site-selection system. In this paper, we design SpoVis, an interactive visual analysis system for planning sports facilities as well as site selection. SpoVis provides users with the distribution status and statistical analysis of various sports facilities. Based on a comprehensive consideration of city population distribution, construction cost, existing sports facilities, traffic situation and development potential, SpoVis provides users with a reasonable site-selection scheme of sports facilities from both macro and micro perspectives and recommends results through topology and map. Meanwhile, based on the distribution of existing sports facilities and city influencing factors, a set of visual analysis components are designed to facilitate users to evaluate the status and information of existing sports facilities. We have carried out extensive experiments on a real platform with real-world data. The experimental results show that the proposed site-selection models and algorithms have excellent accuracy and operation efficiency.**

***Index Terms*—Digital Twinning City, Data Analysis, Sport facility Site Selection, Visual Analysis.**

I. Introduction

DIGITAL twin technology can be widely used in fostering industry development and city informatics by constructing the one-to-one correspondence between the physical city world and the virtual space [1]. Digital twinning cities[1] as a mapping from physical cities to virtual cities can also be regarded as a comprehensive technical solution in the construction of smart cities, considering city planning, city design, service provision, safety, and operational efficiency. Despite the advances of digital twinning technologies in either financial data analysis in digital commerce, city transportation systems or other fields [2]–[5], few efforts have been spent on intelligent site selection of sports facilities. However, it is still one of the vital city planning problems in establishing digital twinning cities with people's lifestyle becoming increasingly diversified.

Most of the previous studies about site selection of sports facilities focus on either data collection, such as obtaining location distribution data of sports facilities, or data analysis with data mining and non-linear optimization tools [6]. However, almost none of them make full use of the technological advantages of digital twinning cities due to the following challenges:

- *Assessment and modeling:* As a bridge between the physical world and virtual world, digital modeling is the basis of implementing digital twin in practice. A critical question is how to build a viable digital twin model for real scenarios [7]. Regarding the site selection, previous studies in various scenarios [6], [8]–[12] indicate that the locations of the sports facilities are mainly affected by population distribution, construction cost, existing sports facilities, traffic situation, and development potential. However, the prior models cannot be directly used in site selection of sports facilities. It is non-trivial to construct suitable mathematical models for assessment and modeling of site selection.
- *Complex data processing and fusion:* It is another challenge to process and fuse massive heterogeneous data collected from various channels such as sensors, physical environment, virtual space, historical database in smart cities [2], [3], [5], [7]. Moreover, the site selection scheme of sports facilities in a city needs to be adaptive to support different application scenarios in digital twinning cities. Thus, the system has critical requirements on both data processing and data fusion.
- *Visualization of results:* An efficient visualization system can help decision-makers propose a better selection. Thus, it is a necessity to investigate how to design a suitable visualization structure. However, visualization has confronted a series of controversial open issues due to the diversity of different fields [13]. Meanwhile, Sedlmair et al. [14] also indicated that a visualization project also needs to analyze a specific real-world problem. Therefore, researchers should conduct specific visualization designs for specific visualization requirements [15]–[21] though it is a challenging task.

To address the challenges above, we propose and develop a Sport facility Visual analysis system (SpoVis) with the provision of the site selection model of sports facilities of the

This work was supported in part by the Joint Key Research and Development Project between Sichuan and Chongqing under Grant cstc2020jscx-cylhX0004, in part by the Macao Science and Technology Development Fund under Macao Funding Scheme for Key R & D Projects under Grant 0025/2019/AKP.

K. Zhang, H. Liu and Z. Lin are with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. (e-mail: kezhang@uestc.edu.cn; hongbo.liu@uestc.edu.cn; 202052081117@std.uestc.edu.cn)

H. Chen is with the Alibaba Group, Hangzhou, China. (e-mail: xunyi.ch@alibaba-inc.com)

H.-N. Dai is with the Department of Computing and Decision Sciences, Lingnan University, Hong Kong SAR, China. (e-mail: hndai@ieee.org)

[1] https://www.challenge.org/insights/digital-twins-and-smart-cities/

city. The major contributions of this paper are summarized as follows:

1) After quantifying the main factors in the site selection of sports facilities, we construct credible mathematical models.
2) We propose optimization algorithms for the proposed macro planning and micro-site selection models to solve the multi-parameter non-linear optimization problem.
3) We devise a set of visual components to facilitate the site selection of sports facilities. The unique interactive analysis of SpoVis makes the results of the site selection more scientific and credible.
4) We design a universal digital-twinning framework for smart city applications. This framework includes a data layer, macro planning engine, micro selecting engine, topological structure engine, and digital twinning layer. The framework is also portable to be applied to different application scenarios. Empirical experiments also demonstrate the effectiveness of the proposed framework.

We organize this paper as follows. Section II presents related work. Section III describes the problems and presents the system framework. Section IV gives the implementation details. Section V evaluate the performance of our proposed models and algorithms based on real data. Finally, the paper concludes in Section VI.

II. Related work

This section introduces related studies being roughly categorized into three aspects: facility-selection algorithms, visual analysis systems, and visualization of site selection.

It is a complex problem to plan and select facilities sites because it is affected by multiple parameters. Hammad et al. [8] proposed a general Mixed Integer Programming (MIP) model for Site Layout Planning Problem (SLPP) of facilities site-selection with the presence of obstacles, formed a spatial discretization scheme, and developed a comprehensive analysis method to inspire evaluation criteria of facilities site-selection results. Jouzdani et al. [9] considered the traffic congestion and the uncertainty of actual demand in the scenario of dynamic facility planning and supply chain design of dairy products and put forward the practical application scheme while this paper emphasizes how to model and apply under the impact of multiple factors. Zhang et al. [10] proposed the Super-Efficiency DEA (SE-DEA) model and the hierarchical node dynamic optimization model based on MIP to solve the common addressing problem of different hierarchical logistics nodes in modern logistics. Ahmad et al. [11] designed a novel data distribution technology after combining with the k shortest path algorithm to re-plan the route. They realized the real-time description of route planning and data distribution in the Internet-of-Things (IoT) scenario. The locations of facilities usually need to be combined with a geographic information system. Esmaelian et al. [12] proposed combining multi-standard decision-making methods with the Geographic Information System (GIS) to connect the location selection criteria of facilities with decision-makers' preferences. Although GIS provides an intuitive and visual window for analysts, the interactivity of the system and visualization of data information are deficient.

Recently, visual analysis systems have been widely combined with many professional fields to provide reasonable and efficient solutions. For example, Karamshuk et al. [15] analyzed user mobility and geographical characteristics, evaluated, and built a layout optimization model based on social network environment data. The SmartAdP visualization system proposed by Liu et al. [16] utilizes large-scale taxi trajectory data to solve the problem of accurate promotion or advertisement delivery conversion rate in the advertising industry. Wu et al. [17] visualized telecom data, proposed TelCoVis to study the co-occurrence of city population flow, and analyzed the co-occurrence model of population flow. The application cases in these professional fields show that the visualization system is efficient and intuitive for data mining. Rathore et al. [18] collected and processed much social network data to monitor diseases, time, public opinion, etc., to support future real-time decision-making and promote plans. Moreover, Din et al. [19] designed a healthcare architecture to collect and analyze health data, which provides valuable information for the healthcare decision-making and management.

The recent technical advances due to rich data sets and interaction forms bring opportunities to design intuitive and flexible interfaces of user interactions using data mining and visual analysis. The facility-location problem strongly correlates with spatio-temporal and multi-dimensional data. Thus, a reasonable and reliable solution can be found with the help of visual analysis. Weng et al. [20] proposed a visualization system of ReACH to help users find, evaluate, and select ideal residences. The system allows users to input multiple criteria, especially the accessibility criteria of residential address and daily activity range that have not been considered in previous solutions to related problems. To provide site-section reference for constructing some facilities, Nivan et al. [21] investigated the taxi activities in different districts based on the trajectory data of taxis and then analyzed people' travel habits. Unlike most traditional visualization systems that use a single data set, Vaud proposed by Chen et al. [22] allows users to conduct joint analysis across multiple data sets to realize the powerful interactive decision exploration.

III. Problem formulation and system framework

A. Data Description and Factor Analysis

Sports facilities have diverse types and various venues, from small fitness equipment in the community to large sports fields. Meanwhile, different sports facilities also have different bearing service rates, service radius, geographic locations, and characteristics. Therefore, the heterogeneity of various sports facilities pose challenges in establishing a site selection model. This paper considers four typical sports facilities defined as $\mathsf{Categories} = \{\mathsf{Equipment}, \mathsf{Fitness\ Centers}, \mathsf{Outdoor\ Courts}, \mathsf{Stadiums}\}$.

We choose Xinwu district in Wuxi City as an example. The data used in this paper come from the city government's public data, web-crawled data, and map supplier. The data can be divided into three types: i) statistical information of

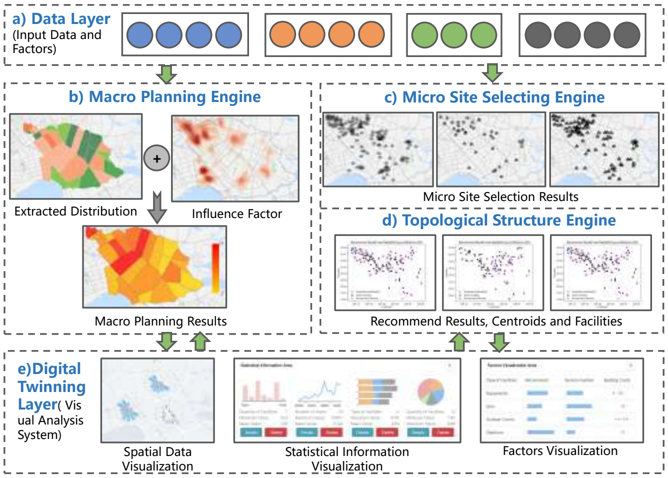


Fig. 1. Framework of SpoVis system

sports venues (such as the number of existing facilities in the city, the investment, and opening time), ii) the geographic locations of various sports facilities (such as the latitude and longitude of the facilities, the community attributions), and iii) the functional district division and Point of Interest (POI) data (such as city commercial areas, residential areas, industrial areas, subway stations, and bus stops). We choose 70% of the city's sports facilities data as the training dataset and the rest 30% of them as the verification set to verify the calculation results. Simultaneously, we collect time-series data of the number of sports facilities in different city districts. We exploit the time-series analysis to predict future trends to determine the total number of different types' sports facilities.

B. Problem Formulation

1) Analysis and statistics of facility data: The analysis and statistics of sports facilities data are conducive to improving system recommendation results and facilitating user's interactive exploration in the visual analysis system via the integration with the data visualization method and the surrounding facilities information.

2) Recommendation of facility site selection: Suppose the city needs M sports facilities to invest. We divide the city into $\mathsf{Districts} = \{d_1, d_2, ..., d_N\}$, where N is the total number of city districts. The macro planning problem is mainly limited by population distribution, construction cost, competition coefficient, traffic convenience, and city development parameter of each district. The goal is to get the best vector $\mathsf{OptNum}_i = \{k_{i,1}, k_{i,2}, ..., k_{i,N}\}$, where $i \in \mathsf{Categories}$ and k is the type i facilities' number in every district. Because each influencing factor's quantitative indicators affect each other, the calculation process of the model is a non-linear optimization problem with multi-parameter constraints.

The derivation of specific locations for facilities is another crucial issue after the number of facilities in each district is obtained. Based on the population distribution division, requirement centers of various sports facilities can be calculated through a clustering algorithm. The cost time between the recommended location k of a sport facility and one of the requirement centers l of type i facilities in district j is $t_{i,j}^{k,l}$. The distance between the recommended location of a facility and


(a) Intensive population places: office buildings (blue), schools (green), communities (yellow), and dormitories (red)
(b) Heat map of population density (darker color means higher population)

Fig. 2. Population density estimation

an existing facility m of type i in district j is $d_{i,j}^{k,m}$. We then build a model to find the optimal locations where the values of these two variables are as small as possible. Because the time needed to get to the requirement centers and the distance to the existing facilities changes with the varied recommended locations, the calculation of specific recommended locations is a non-linear optimization problem.

C. SpoVis Framework

Fig. 1 depicts the SpoVis framework including five components: a) *Data Layer (Input Data and Factors)*; b) *Macro Planning Engine*; c) *Micro Site Selecting Engine*; d) *Topological Structure Engine*; e) *Digital Twinning Layer (Visual Analysis System)*. We next present the detailed design of SpoVis.

IV. System Architecture

A. Macro Planning

We divide the city into multiple districts according to communities and main transportation arteries. We then use population-intensive POI data to obtain the city's population distribution. The fusion of them is conducive to establish a normalized variable model.

1) Population Distribution Extraction: Some recent studies use telecommunication data to analyze the flow and distribution of city population. In contrast, the difficulty of obtaining telecommunication data is not conducive to expanding and promoting the system. Since map suppliers own rich data resources, we can effectively get the locations of intensive population places like office buildings, schools, communities, and dormitories through POI search. We use the obtained locations data to estimate the population's spatial distribution. Fig. 2 shows final results.

2) Districts Division: If considering main transportation arteries of a city only, the sole division of the city area will ignore the attributes of the functional areas in the city. As shown in Fig. 3, we use the city's arterial transportation roads as districts' boundaries and initially divide the city into multiple communities. We next calculate the similarities of every two communities. The features of communities include five influence factors of macro planning, as mentioned in Section III. We denote the similarities between community A and B by $\mathsf{Similarity}(A, B)$. Specifically, the features of community A are represented by a 5-dimensional variable

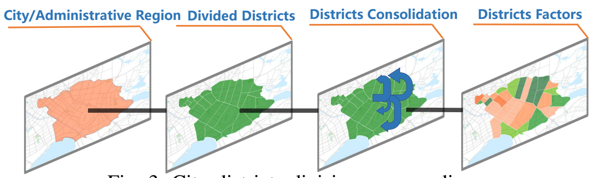


Fig. 3. City districts division process diagram

$A(x_1, x_2, ..., x_5)$, and the features of community B are represented by $B(y_1, y_2, ..., y_5)$ similarly. Then $\mathsf{Similarity}(A, B)$ can be calculated by the following equation:

$$\mathsf{Similarity}(A, B) = (\sum_{i=1}^{5} w_i |x_i - y_i|^2)^{\frac{1}{2}}, \tag{1}$$

where the weight of feature i is denoted by w_i. We then have $\sum_{i=1}^{5} w_i = 1$.

3) Multi-parameter Normalization: In site-selection model of the sports facilities, many factors need to be quantified and normalized to avoid the deviation of the model conclusion caused the data difference of certain factors. The factors to be quantified and normalized include the following:

(a) *Population density*: We use POI-searching API provided by the map supplier to get the locations of some population-intensive places like office buildings, dormitories, schools, and communities. We then use these locations to estimate the population density of districts. The population density of district j (where $j \in \mathsf{Districts}$), is $x_j^{\text{density}} = \sum_{i=1}^{4} h_i \cdot n_{ij}$, where h_i is the average number of floors of type i building, representing the height of the building; n_{ij} denotes the number of type i building in district j; and the value of i is from 1 to 4, separately representing one of the four intensive population places. Similarly, the city's total population density is $x_D^{\text{density}} = \frac{\sum_{j=1}^{N} x_j^{\text{density}}}{N}$, where N is the number of Districts. Then the normalized population density in district j is $F_j^{\text{density}} = \frac{x_j^{\text{density}}}{x_D^{\text{density}}}$.

(b) *Construction cost*: The normalized construction cost of a specific sports facility i ($i \in \mathsf{Categories}$ in district j) can be obtained by $F_{i,j}^{\text{cost}} = F_{i,j}^{\text{coverage}} \cdot \frac{\mathsf{Cost}(n_j^{\text{business}}, n_j^{\text{living}}, n_j^{\text{school}})}{\mathsf{Cost}(n_D^{\text{business}}, n_D^{\text{living}}, n_D^{\text{school}})}$, where $F_{i,j}^{\text{coverage}}$ denotes the coverage area of type i facilities in district j; $\mathsf{Cost}(\cdot)$ is a counting operation of parameters; terms n_j^{business}, n_j^{living}, and n_j^{school} indicate the number of business centers, residential centers, and schools in district j, respectively, while the n_D^{business}, n_D^{living}, and n_D^{school} indicate the numbers of these buildings in the whole city, respectively. They significantly affect nearby housing prices.

(c) *Competition coefficient*: The competition coefficient is calculated by mapping the number of existing facilities denoted by $N_{i,j}^{\text{exist}}$ ($i \in \mathsf{Categories}$ and $j \in \mathsf{Districts}$) to the population density of each district and that of the whole city denoted by F_j^{density} and F_D^{density}, respectively. The normalized competition coefficient is $F_{i,j}^{\text{competitive}} = \frac{N_{i,j}^{\text{exist}}}{F_j^{\text{density}}} / \frac{N_{i,D}^{\text{exist}}}{F_D^{\text{density}}}$, where $F_D^{\text{density}} = 1$ represents the relative population density of the whole city.

(d) *Traffic convenience*: The number of bus stops and the number of subway stations in district j are denoted by n_j^{bus}, n_j^{subway}, respectively. Similarly, the number of bus stops and the number of subway stations in the whole city are denoted by n_D^{bus}, and n_D^{subway}, respectively. The normalized traffic factor is $F_j^{\text{traffic}} = \frac{N_j(n_j^{\text{bus}}, n_j^{\text{subway}})}{N_D(n_D^{\text{bus}}, n_D^{\text{subway}})}$, where N_j and N_D are counting operations.

(e) *Development parameter*: As a long-term public infrastructure serving the masses, locations of sports facilities need to satisfy the spatial distribution of current demand and the development potential of each district. The district that has the potential to attract residents in the future will show tremendous competitive advantages. The district development parameter defined according to the following principles: i) the closer to the subway station, the greater competitiveness is; ii) the larger number of industrial and commercial areas is, the greater competitiveness is; iii) governments, hospitals, schools and other social infrastructures are developed regionally, the greater competitiveness is. Therefore, we define the development parameter of district j as $D_j^{\text{dev}} = n_j^{\text{school}} + n_j^{\text{subway}} + n_j^{\text{hospital}} + n_j^{\text{government}} + n_j^{\text{market}} + n_j^{\text{factory}}$ and the development parameter of the whole city as $D_D^{\text{dev}} = n_D^{\text{school}} + n_D^{\text{subway}} + n_D^{\text{hospital}} + n_D^{\text{government}} + n_D^{\text{market}} + n_D^{\text{factory}}$. Finally, we can get the relative development parameter of each district by normalization $F_j^{\text{dev}} = \frac{D_j^{\text{dev}}}{D_D^{\text{dev}}}$.

4) Macro Planning Model: We denote the best-recommended number of type i sports facilities in district j by $k_{i,j}$. We define the total number of type i facilities to invest as $\mathsf{TotalNum}_i = \sum_{j=1}^{\mathsf{Districts}} k_{i,j}$ as a constraint condition, where $k_{i,j} \in \mathsf{OptNum}_i$. We then combine the above-mentioned factors and the number of facilities as parameters to construct the following objective function:

(a) *Population density*: Due to the apparent linear correlation between population density and the number of facilities needed, the parameter of population density denoted by P_{density} can be calculated by $P_{\text{density}} = \sum_{j=1}^{\mathsf{Districts}} (F_j^{\text{density}} k_{i,j})$, where F_j^{density} is the normalized city population density of district j.

(b) *Construction cost*: The construction cost of sports facilities is affected by the facilities' construction area, price per unit area and other factors, thereby having a non-linear relationship with the number of facilities in a district. So, the parameter of city construction cost denoted by P_{cost} can be expressed as $P_{\text{cost}} = \sum_{j=1}^{\mathsf{Districts}} (F_{i,j}^{\text{cost}} e^{-k_{i,j}})$, where $F_{i,j}^{\text{cost}}$ is the construction cost of type i sports facilities in district j.

(c) *Competition coefficient*: The existing sports facilities have a great influence on the macro planning results and the interference process is mostly nonlinear. Therefore, the competition coefficient parameter of the same type of sports facilities is expressed as $P_{\text{competitive}} = \sum_{j=1}^{\mathsf{Districts}} (F_{i,j}^{\text{competitive}} e^{-k_{i,j}})$, where $F_{i,j}^{\text{competitive}}$ is the competition coefficient of type i facilities in district j.

(d) *Traffic convenience*: The traffic influence on medium and large sports facilities plays a key role in location decision while it varies within a specific range. We thus evaluate the traffic convenience, which is expressed as

$P_{\text{traffic}} = \sum_{j=1}^{\text{Districts}} (F_j^{\text{traffic}} \frac{1}{1+e^{-k_{i,j}}})$, where F_j^{traffic} is the traffic convenience of district j.

(e) *Development parameter*: The development parameters reflect the development trend of population distribution in the future, thereby essentially serving as the response of population density in the future. Therefore, it has a linear relationship with the number of facilities in each district. We define it as $P_{\text{dev}} = \sum_{j=1}^{\text{Districts}} (F_j^{\text{dev}} k_{i,j})$, where F_j^{dev} is the development parameter of the district j.

The objective function of type i facilities of the optimization model is defined as follows:

$$\begin{aligned} \mathsf{RMacro}_i = \max(k_{i,j}) = w_1 P_{\text{density}} - w_2 P_{\text{cost}} - \\ w_3 P_{\text{competitive}} + w_4 P_{\text{traffic}} + w_5 P_{\text{dev}}, \end{aligned} \tag{2}$$

where the weight of each parameter w_i is adjustable to fit different application scenarios and comply with the constraint condition $\sum_{i=1}^{5} w_i = 1$; and $k_{i,j}$ must meet $\sum_{j=1}^{\text{Districts}} k_{i,j} = \mathsf{TotalNum}_i$.

5) Algorithm Optimization: The above non-linear programming problem involves the optimization of multi-parameter functions. Genetic Algorithms (GAs) [23], [24] have achieved excellent performance in the global search scenario due to the usage of selection, crossover, and traversal operators to search for the optimal solution. However, the performance of them may not be scalable when the number of parameters increases, especially in some specific application scenarios.

Traditional GAs have two critical parameters: crossover rate P_c and mutation rate P_m while the choice of these two parameters also has a certain ambiguity. To address this problem, we devise an Improved adaptive GA (IGA) to adjust P_c and P_m during the entire optimization process as follows:

$$P_c = \begin{cases} k_1(f_{\max} - f')/(f_{\max} - f_{\text{ave}}), f' \geq f_{\text{ave}} \\ k_2, f' < f_{\text{ave}} \end{cases}, \tag{3}$$

$$P_m = \begin{cases} k_3(f_{\max} - f)/(f_{\max} - f_{\text{ave}}), f \geq f_{\text{ave}} \\ k_4, f < f_{\text{ave}} \end{cases}, \tag{4}$$

where $f_{\max}$ is the maximum fitness of the population, f_{ave} is the average population fitness, f is the fitness value of the mutant individual, f' is the adjusted fitness value between two individuals (the better one), k is the adjustable coefficient.

In order to make the crossover rate and mutation rate non-linearly be adapted to the optimization process according to the degree of population fitness dispersion, we propose a non-linear method to adjust P_c and P_m, to ensure the evolution of high-quality population individuals, accelerate the convergence speed, and search for the optimal global solution. Accordingly, P_c and P_m can be calculated as follows:

$$P_c = \begin{cases} k_1 \frac{\arcsin(f_{\text{ave}}/f_{\max})}{\pi/2}, \arcsin(f_{\text{ave}}/f_{\min}) < \frac{\pi}{6} \\ k_1(1 - \frac{\arcsin(f_{\text{ave}}/f_{\max})}{\pi/2}), \arcsin(f_{\text{ave}}/f_{\min}) \geq \frac{\pi}{6} \end{cases}, \tag{5}$$

$$P_m = \begin{cases} k_2(1 - \frac{\arcsin(f_{\text{ave}}/f_{\max})}{\pi/2}), \arcsin(f_{\text{ave}}/f_{\min}) < \frac{\pi}{6} \\ k_2 \frac{\arcsin(f_{\text{ave}}/f_{\max})}{\pi/2}, \arcsin(f_{\text{ave}}/f_{\min}) \geq \frac{\pi}{6} \end{cases}. \tag{6}$$

It is worth mentioning that we introduce the degree of dispersion $\arcsin(f_{\text{ave}}/f_{\max})$ of the population fitness, and adjust the order of crossover and mutation according to the degree of dispersion, thereby fastening the operation speed.


(a) Stadiums centroids (Total No. = 30)


(b) Outdoor courts centroids (Total No. = 40)


(c) Fitness centers centroids (Total No. = 100)

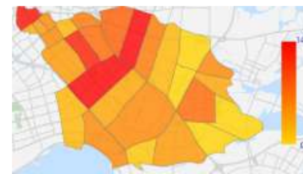

(d) Equipment centroids (Total No. = 200)

Fig. 4. Macro planning results for each type facilities

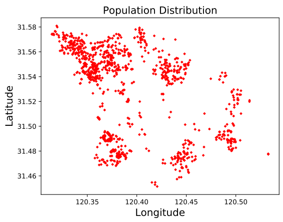
After putting the objective function and constraints of the macro planning model into our algorithm, we finally get the macro planning results, as shown in Fig. 4.

B. Micro Site Selection

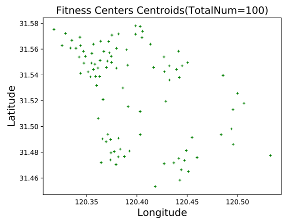
To maximize the usage of recommended sports facilities in every district, we establish a micro-site selection model. The construction of the micro-site selection model includes the following steps: 1) obtaining the cluster centers of intensive population POI data through the hierarchical spatial clustering method; 2) normalizing the parameters of the micro-site selection model; 3) using an improved Simulated Annealing Algorithm (SAA) [24] to calculate the specific recommended locations in the district; 4) introducing the network topology figures to represent the distribution relationship between the recommended results of sports facilities as well as cluster centers and existing facilities separately; 5) visualizing the final results on the map. These processes help the comparison and adjustment of model parameters.

1) Requirement centers extraction: As shown above, we obtain the locations of intensive population places in the whole city. Because sports facilities ultimately provide people with services, it is necessary to calculate the population requirement centers based on the gotten POI locations data. We use the hierarchical spatial clustering method to calculate the cluster centers, i.e., $\mathsf{Centers}_i^{\text{living}} = \{C_{k_{i,1}}, C_{k_{i,2}}, ..., C_{k_{i,N}}\}$, where $C_{k_{i,j}}$ represents the centers of type i sports facilities in district j and the number of centers of the type i facilities in the city is $\mathsf{TotalNum}_i$. Theoretically, these centers could be used as the initial locations of the algorithm while this method has its own shortcoming. These centers will serve to construct the optimization model. The clustering results are shown in Fig. 5.

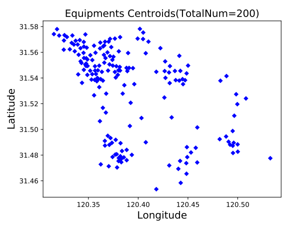
2) Parameter normalization and modeling: The factors that affect the precise locations of sports facilities can be summarized as follows: the weighted average cost time related to bicycles, public transportation, and private transportation, the distance from the specified recommended locations to the cluster centers defined above, and the distance from the specified recommended locations to the existing facilities



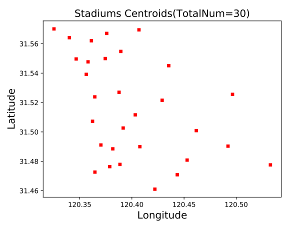
(a) Population distribution



(b) Fitness centers centroids (Total No. = 100)



(c) Equipment centroids (Total No. = 200)

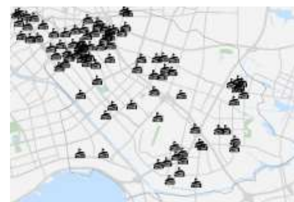


(d) Stadiums centroids (Total No. = 30)

Fig. 5. Clustering distribution results as the demand centers of sports facilities

(a) Stadiums centroids (Total No. = 30)

(b) Outdoor courts centroids (Total No. = 40)



(c) Fitness centers centroids (Total No. = 100)

(d) Equipment centroids (Total No. = 200)

Fig. 6. Micro site selection results for each type of facilities

nearby. Therefore, when constructing our micro site selection model, the main parameters can be normalized as follows:

(a) We denote the recommended location k by $L_{i,j}^{k} = \mathsf{Loc}(x_k^{\text{rec}}, y_k^{\text{rec}})$. In district j, the weighted average time cost of type i facilities from location k to all the cluster centers is denoted by $t_{i,j}^{k} = \sum_{l=1}^{n} t_{i,j}^{k,l}$, where n is the total number of cluster centers in district j and $t_{i,j}^{k,l}$ represents the weighted average time cost of recommended location k of type i in district j to cluster center l, which is from the path planning function provided by the map supplier. We fully consider travel modes, including the buses, self-driving, bicycles, and subways. We weigh average the four types of travel time, and finally get the commuting expenses between the recommended location k and the cluster center l.

(b) The distance from the recommended location k to the existing facility m of type i in district j is denoted by $d_{i,j}^{k,m}$. The distribution of existing facilities within the service range of the recommended facility influences the micro site selection results, and the distance between the location $L_{i,j}^{m} = \mathsf{Loc}(x_m^{\text{exist}}, y_m^{\text{exist}})$ of existing sports facilities and the recommended location in district j of type i is expressed as $d_{i,j}^{k,m} = \sqrt{(x_k^{\text{rec}} - x_m^{\text{exist}})^2 + (y_k^{\text{rec}} - y_m^{\text{exist}})^2}$. So, the sum of the distances between a recommended location k and all other existing facilities in its district is given as $d_{i,j}^{k} = \sum_{m=1}^{n} d_{i,j}^{k,m}$, where n is the number of existing type i facilities in district j.

After calculating each normalized parameter, the number of type i sports facilities planned to be built in district j is fixed under the constraint $k_{i,j}$. We then obtain the micro site selection results denoted by $\mathsf{RMicro_i}$ when the number of the objective function denoted by $\min(L_{i,j}^{k})$ is minimum. In particular, we have

$$\mathsf{RMicro_i} = \min(L_{i,j}^{k}) = \sum_{j=1}^{\text{Districts}} w_1 t_{i,j}^{\text{k}} - w_2 d_{i,j}^{\text{k}}, \tag{7}$$

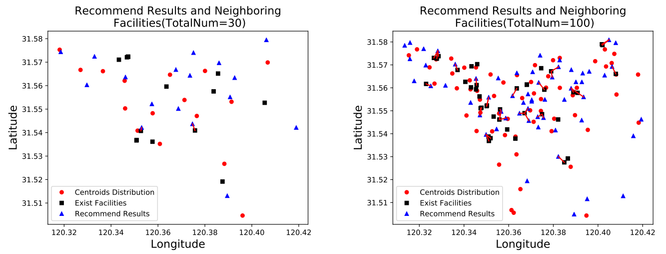
where $w_1 + w_2 = 1$.

3) Algorithm optimization: According to the site selection model above, the system uses the spatial distribution distance and travel time as factors for function optimization. Taking the population distribution centers, which were mentioned in the Requirement center extraction section as the initial population, the optimization process can easily fall into the local optimum, although the result can be calculated quickly. As a heuristic algorithm, the SAA calculates the new solution's objective function deviation and judges whether it can replace the current feasible solution according to the Metropolis criterion.

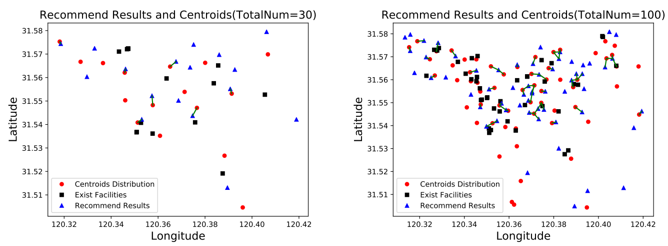
Therefore, we adopt a hybrid method to select a random population and adjacent points as the initial population and use the improved GA to conduct global optimization calculations. The local search with SAA costs less time to obtain better and more accurate recommendation results. Our proposed Hybrid Genetic Simulated Annealing (HGSA) method integrates both the time efficiency and accuracy of the above two algorithms to achieve better results in the scene of sports facilities site selection. We finally get the results, as shown in Fig. 6.

4) Topological structure: The topological structure is used to abstract network structure to obtain more intuitive comparison results. Therefore, we construct the topological structure by mapping the recommended results, the locations of existing sports facilities, and city residential points to the topology structure nodes. Then, we connect the nodes according to the service radius of each type of sports facilities.

As shown in Fig. 7, the black squares indicate the existing facilities in the area, and the red dots indicate the residential spots or resident intensive POI distribution, and the blue triangles indicate the recommended facilities. These nodes indicate their spatial distribution. Once the nodes' conflicts occur in the distribution radius, the two conflicted nodes will be connected by an edge. In Fig. 7(a), when there are existing sports facilities (i.e., nodes) within the recommended facilities' service radius, the two nodes are connected with red lines. The more complex the topology is, the less ideal the recommended results are. In Fig. 7(b), when there are



(a) Recommend results and neighboring facilities



(b) Recommend results and centroids

Fig. 7. Topological structures of facilities site-selection results

intensive population POI nodes within the service radius of the recommended facilities, they are connected with green lines. The more complex the topology is, the more ideal the recommended results are.

C. Interactive Visual Analysis System

The results of macro planning and micro-site selection of sports facilities need to be presented in combination with geographic information [25]. Due to the high dimension of data, it is quite complex to display the information. The interactive visual analysis system can show users rich statistical information and internal correlation of data [26], [27]. Thus, we propose an interactive visual analysis system as shown in Fig. 8, which mainly includes the following components:

1) Interactive control bar, in component A of Fig. 8: this component is used to select the city and import relevant statistical data. The user can adjust the macro planning and micro-site selection model weights through the interactive control bar to get the most suitable recommendation model.

2) Spatial information visualization, in component B of Fig. 8: the city population distribution is one of the critical elements of sports facilities recommendation scheme. We have obtained the population distribution and its corresponding heat map, both of which can be used for the decision-makers to adjust the city planning parameters.

Recommendation model results: Our site-recommendation scheme mainly consider macro planning and micro-site selection, both of which can be visually displayed in our system. In the macro planning, different colors represent the varied number of facilities of districts and the dark color implies more facilities. In the micro-site selection model, we show the final recommended locations of the facilities on the map.

Regional population flow statistics: The flow of information of people in different areas of the city profoundly impacts sports facilities' location decisions. We use the traffic data of the city's main areas to count the crowd flow. Meanwhile, we

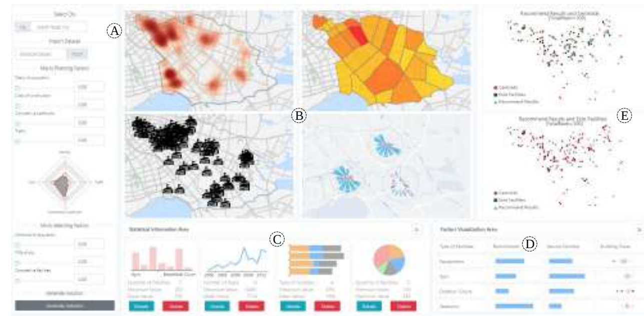


Fig. 8. Interactive visual analysis system

adopt the polar charts to show the distribution characteristics of the data. As shown in Fig. 9, we count the population within 12 hours and show the statistical values in each interval through the cascade histogram. With the chart, analysts can quickly capture the variation of crowd flow in the region.

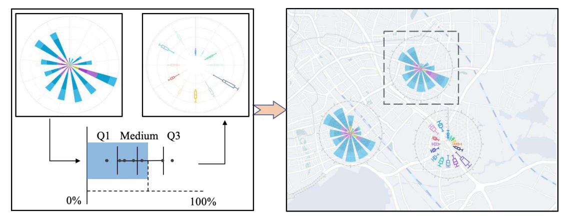


Fig. 9. Statistical charts of city population flow

3) Visualization of statistical information, in component C of Fig. 8: the statistics of sports facilities data is a vital component of the site recommendation strategy. With the increased types of sports facilities, sports facilities data has various dimensions, such as construction number, construction time, and score. We exploit the multi-view collaboration to display the statistical data of city sports facilities through multiple layers and panes. To enable users to explore the data of specific facilities more efficiently and intuitively, we construct associations between charts. Users can realize the linkage between charts by swiping and highlighting to achieve the exploration effect. The basis of the multi-view collaboration is the statistic, sorting, and combination of raw data. Fig. 10 shows the schematic diagram of the method of data statistic and reorganization. This method can efficiently organize data and complete data filtering and calculation under different conditions. Interactive control bar in component A of Fig. 8 can set data combination and rendering of charts. When multiple measures are displayed in the chart, the measurement unit is non-uniform. It needs to use the measure name (i.e., M_{name}) and measure value (i.e., M_{value}) as dimensions and measures to generate the chart. The process can be expressed as follows:

$$\mathsf{data, measure} \rightarrow \sum_{i=0}^{i<\text{measures.length}} \mathsf{data} \bigcup \mathsf{Measure}_{\text{expond}}, \quad (8)$$

$$\mathsf{Measure}_{\text{expond}} = \{(m_j, v_j) | m_j \in M_{\text{name}}, v_j \in M_{\text{value}}\}. \quad (9)$$

4) Comparison of model parameters in component D of Fig. 8: the recommendation model of sports facilities includes multiple parameters and weights. The varied parameters also affect site recommendation. Due to the significant differences

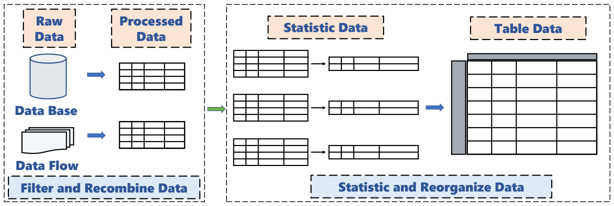


Fig. 10. Schematic diagram of the data statistic and reorganization

between cities, the adjustable parameter enhances the versatility of the site recommendation system. Users can conduct multiple tests according to different parameter combinations to obtain a good practice scheme. As shown in component D of Fig. 8, we normalize the factors for the site recommendation model. For a single type of value, we use the column form to compare different sports facilities. For the factors composed of multiple data points, such as construction cost, we use a box line chart to show the data distribution characteristics.

5) Topology validation, in component E of Fig. 8: the micro-location results of facilities need to be combined with GIS and can be observed and verified through the topology. We use network nodes to represent the distribution of recommendation results, existing facilities, and population cluster centers. Meanwhile, we use connection lines to represent their coverage relationship. It is helpful for users to intuitively and efficiently compare the results.

V. Experiments

A. Experimental Settings

1) Baseline algorithms: Both macro planning and micro-site selection are nonlinear multi-parameter optimizations. Since they have different parameter settings and accuracy requirements, we select different baselines for each problem.

- Macro planning: Adam algorithm, Ant Colony Algorithm (ACA), and GA.
- Micro-site selection: Nadam algorithm, GA and SAA.

2) Evaluation metrics: We adopt different evaluation metrics to evaluate algorithms in each of the above problems. In particular, we have the following evaluation metrics.

- Accuracy of macro planning: The macro planning is a multi-parameter nonlinear optimization problem. We adopt the accuracy to evaluate the performance of algorithms for macro the planning as follows,

$$\mathsf{Accuracy}_{i,j} = 1 - \frac{\sum_{D}^{j=1} |y_{i,j} - \widehat{y_{i,j}}|}{N_i}, \quad (10)$$

  where there are D districts in a city, the recommended number of sports facilities i for the district j is denoted by $y_{i,j}$, the actual number of constructions invested is $\widehat{y_{i,j}}$, and the total number of constructions invested is N_i.
- Evaluation of micro-site selection: For k_{th} $\big(k \in \{1,2,...,K_{i,j}\}\big)$ recommended facility in district j, the spatial location is $L_k^{\text{rec}} = \mathsf{Loc}(x_k^{\text{rec}}, y_k^{\text{rec}})$, and its real location is $\hat{L_k} = \mathsf{Loc}(\widehat{x_k}, \widehat{y_k})$. The deviation distance is evaluated by ErrorDistance given as follows,

$$\mathsf{ErrorDistance}_{i,j} = \sum_{k=1}^{K_{i,j}} (\sqrt{(x_k^{\text{rec}} - \widehat{x_k})^2 + (y_k^{\text{rec}} - \widehat{y_k})^2}). \quad (11)$$

(a) Rasterizing the city, where each small rectangle represents an area (b) The darker each area implies the easier to obtain facilities and services

Fig. 11. Spatial accessibility

  The larger value of ErrorDistance implies the more deviation between the recommended results and the real data (i.e., the poorer performance of an algorithm).

3) Facilities spatial accessibility and equity : We choose a widely-used spatial accessibility measurement named the potential model [28], which consists of three metrics: the availability of facilities to provide services of area i denoted by A_i, the impact of the population of all the m areas on facility j denoted by V_j, the capacity of facility j on area i denoted by S_{ij}. We then give the expressions of A_i, V_j, S_{ij} as follows,

$$A_i = \sum_{j=1}^{n} \frac{S_{ij} M_j}{D_{ij}^{\beta} V_j}, V_j = \sum_{k=1}^{m} \frac{S_{kj} P_k}{D_{kj}^{\beta}}, S_{ij} = 1 - (\frac{D_{ij}}{D_j})^{\beta}, \quad (12)$$

where n represents the total number of facilities in the city, D_{ij} is the distance between facility j and area i, P_k represents the population of area k, and β represents the travel friction coefficient that varies with demographic characteristics, types of facilities, services, and other factors.

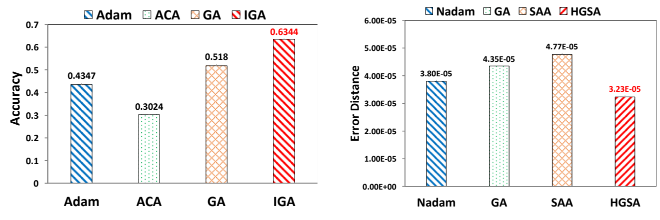
A larger value of A_i implies that the people living in area i have higher chance to obtain the services provided by facilities. In order to calculate A_i, we also need to calculate the carrying rate M_j of facility j, representing the population that facility j can support at the same time. Meanwhile, the capacity S_{ij} depends on D_{ij} and the service radius D_j of facility j.

B. Experimental Results

We evaluate the proposed algorithm and baselines via extensive experiments on a real-world dataset. We compare the performance of algorithms in terms of both macro planning and micro selection. Moreover, we also investigate the impacts of various parameters on the proposed model and analyze parameter configurations in different scenarios. Further, we conduct the experiments to evaluate Spatial Accessibility.

1) Spatial accessibility results: We divide the whole Xinwu district into a number of rectangles, each of which represents an area, as shown in Fig. 11(a). We use the number of intensive population buildings in an area to represent the population. Fig. 11 shows the experimental results. Compared with previous population density estimation in Fig. 2(b), it is not difficult to find that the densely populated areas can get better services (dark red areas), as shown in Fig. 11(b); this further proves the effectiveness of our scheme.

2) Accuracy comparison of different models: We evaluate the accuracy of the proposed algorithms and baseline algorithms on the macro planning and the micro-site selection. Fig. 12 shows a performance comparison of the proposed



(a) Accuracy (b) ErrorDistance

Fig. 12. Accuracy comparison of different algorithms

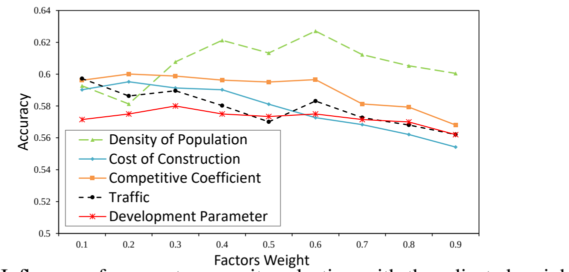


Fig. 14. Influence of parameters on site selection with the adjusted weight

algorithms and baseline algorithms in terms of *Accuracy* and ErrorDistance. In particular, we have the following findings.

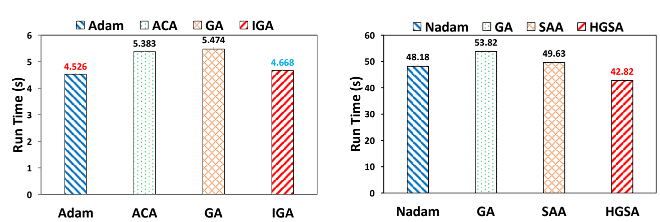
Macro planning. We observe from Fig. 12(a) that our proposed IGA has achieved the highest accuracy (i.e., 0.6344) among all the algorithms. The main improvement of IGA may owe to the adjusted crossover rate and mutation rate so that IGA can quickly obtain the optimal global solution.

Micro-site selection. Fig. 12(b) shows that HGSA has the lowest ErrorDistance value (i.e., 3.23×10e-5) among all the algorithms. The reason can be explained as follows: i) HGSA is superior in selecting random population location and clustering center as initial population location; ii) HGSA can then adopt a heuristic algorithm to obtain optimal results since it is more likely to get the optimal global solution; iii) the space distance between the recommended result and the real location in HGSA is the smallest.

3) Comparison of running time of different models: We also evaluate running time of the proposed algorithms and other algorithms. Similarly, our evaluations can also be categorized into macro planning and micro-site selection. Fig. 13 shows the comparison results. We have the following findings.

Macro planning. As shown in Fig. 13(a), Adam achieves the lowest running time (i.e., 4.526s) among all the algorithms, while our proposed IGA also achieves excellent performance (i.e., 4.668s, only slightly higher than that of Adam). Whereas our IGA has much higher accuracy in macro planning than Adam (as shown in Fig. 12(a)).

Micro-site selection. As shown in Fig. 13(b), our proposed HGSA achieves the lowest running time (i.e., 42.82s) among all the algorithms. The main reason for the superior performance of HGSA may owe to the hybrid method of HGSA, which is used to set the initial population and to adjust the number of iterative calculations. On the premise of avoiding falling into the local optimum, HGSA can further improve the calculation efficiency compared with the Nadam, GA, and SAA algorithms.



(a) Macro planning (b) Micro-site selection

Fig. 13. Running time of different algorithms

4) Comparison of parameters: In the planning of city sports facilities in each district, the weight of each parameter may affect the accuracy of the site selection results. In order to evaluate the influence of each parameter, we fix a weight for one parameter while equally dividing the remaining weight to other parameters. In particular, we consider the following parameters: i) *Density of Population*, ii) *Cost of Construction*, iii) *Competitive Coefficient*, iv) *Traffic*, and v) *Development Parameter*. Fig. 14 shows the comparison results.

As shown in Fig. 14, we observe that the density of the population is an essential factor. When its weight increases (i.e., it plays a more important role), the accuracy shows an overall upward trend (before its weight reaches 0.6). However, the accuracy decreases when its weight goes beyond 0.6. It implies that the weight of the population density should be well configured within a reasonable range (e.g., $[0, 0.6]$). The similar observations can be found for other parameters. In addition, we find that the accuracy tends to decrease with the increase of other parameters.

5) Practical verification in other cities: Throughout this work, we use Xinwu district as a typical example to demonstrate our experimental process and results. To show the portability of our scheme, we supplement an experiment based on Shuangliu district of Chengdu city. The experimental results are shown in Fig. 15. It is not difficult to find that our scheme still has a good performance and spatial accessibility for densely populated areas.

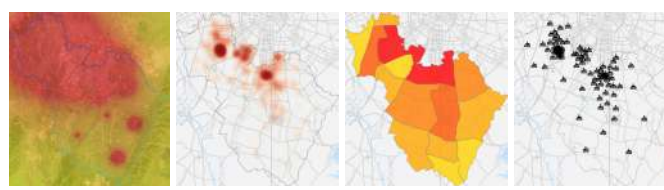


Fig. 15. Experimental results of Shuangliu district

VI. Conclusion

As a city urban service system, the digital twinning system typically includes the following components: a) providing digital modeling as a bridge between the physical world and virtual world; b) obtaining the dynamic information from the physical world, being superimposed on the digital model to achieve a more realistic and intuitive visualization; c) using algorithms and big data systems to process and analyze the acquired data, thereby making corresponding decisions; d) making decisions to the physical world to guide people's behaviors and program implementation after being guided by the digital-twinning system. Based on the above considerations, in this paper, we first quantify the primary factors and several integrated factors in the site selection of sports facilities and construct related mathematical models. Secondly, we propose the optimization algorithms for the proposed macro planning and micro-site selection models to solve the multi-parameter nonlinear optimization problem. Finally, we design the SpoVis system, which can visually display and analyze the real city data, and propose several decision-making algorithms for facility location to help decision-makers implement their scheme. In this step, we also devise a set of visual components. As a result, the unique interactive analysis of SpoVis makes the site selection recommendation more reasonable and reliable. Our empirical results show that both the proposed site-selection models and algorithms have achieved superior performance than other state-of-the-art baseline approaches in terms of high accuracy and short running time.

REFERENCES

[1] C. Gehrmann and M. Gunnarsson, "A digital twin based industrial automation and control system security architecture," *IEEE Transactions on Industrial and Informatics*, vol. 16, no. 1, pp. 669–680, 2020.

[2] T.-M. Choi, H. K. Chan, and X. Yue, "Recent development in big data analytics for business operations and risk management," *IEEE transactions on cybernetics*, vol. 47, no. 1, pp. 81–92, 2016.

[3] X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang, and L. Yang, "Big data for social transportation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 3, pp. 620–630, 2015.

[4] X. Wang, L. Yang, Y. Wang, X. Liu, Q. Zhang, and M. Deen, "A distributed tensor-train decomposition method for cyber-physical-social services," *ACM Transactions on Cyber-Physical Systems*, vol. 3, pp. 1–15, 2019.

[5] X. Wang, L. Yang, X. Chen, M. Deen, and J. Jin, "Improved Multi-order Distributed HOSVD with its Incremental Computing for Smart City Services," *IEEE Transactions on Sustainable Computing*, vol. PP, pp. 1–14, 2018.

[6] F. Hu, S. Yang, and W. Xu, "A non-dominated sorting genetic algorithm for the location and districting planning of earthquake shelters," *International Journal of Geographical Information Science*, vol. 28, no. 7, pp. 1482–1501, 2014.

[7] F. Tao, H. Zhang, A. Liu, and A.Y.C.Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 2405–2415, 2019.

[8] A. W. Hammad, D. Rey, and A. Akbarnezhad, "A cutting plane algorithm for the site layout planning problem with travel barriers," *Computers & Operations Research*, vol. 82, pp. 36–51, 2017.

[9] J. Jouzdani, S. J. Sadjadi, and M. Fathian, "Dynamic dairy facility location and supply chain planning under traffic congestion and demand uncertainty: A case study of tehran," *Applied Mathematical Modelling*, vol. 37, no. 18-19, pp. 8467–8483, 2013.

[10] X.-D. Zhang, H. Pan, and B. Han, "Study on location planning with grade nodes of urban logistics," in *2010 8th International Conference on Supply Chain Management and Information*. IEEE, 2010, pp. 1–7.

[11] A. Ahmad, S. Din, A. Paul, G. Jeon, M. Aloqaily, and M. Ahmad, "Real-time route planning and data dissemination for urban scenarios using the internet of things," *IEEE Wireless Communications*, vol. 26, no. 6, pp. 50–55, 2019.

[12] M. Esmaelian, M. Tavana, F. J. Santos Arteaga, and S. Mohammadi, "A multicriteria spatial decision support system for solving emergency service station location problems," *International Journal of Geographical Information Science*, vol. 29, no. 7, pp. 1187–1213, 2015.

[13] M. Meyer and J. Dykes, "Criteria for rigor in visualization design study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, pp. 87–97, 2020.

[14] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, pp. 2431–2440, 2012.

[15] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geo-spotting: mining online location-based services for optimal retail store placement," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 793–801.

[16] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu, "Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 1–10, 2016.

[17] W. Wu, J. Xu, H. Zeng, Y. Zheng, H. Qu, B. Ni, M. Yuan, and L. M. Ni, "Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 935–944, 2015.

[18] M. M. Rathore, A. Paul, A. Ahmad, M. Imran, and M. Guizani, "Big data analytics of geosocial media for planning and real-time decisions," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[19] S. Din and A. Paul, "Smart health monitoring and management system: Toward autonomous wearable sensing for internet of things using big data analytics," *Future Generation Computer Systems*, vol. 91, pp. 611–619, 2019.

[20] D. Weng, H. Zhu, J. Bao, Y. Zheng, and Y. Wu, "Homefinder revisited: Finding ideal homes with reachability-centric multi-criteria decision making," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.

[21] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2149–2158, 2013.

[22] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan, and R. Maciejewski, "Vaud: A visual analysis approach for exploring spatio-temporal urban data," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 9, pp. 2636–2648, 2017.

[23] S. Odofin, Z. Gao, and K. Sun, "Robust fault estimation in wind turbine systems using ga optimisation," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*. IEEE, 2015, pp. 580–585.

[24] K. Zhang, D. Chang, and H. Jia, "Genetic simulated annealing-based coverage-enhancing algorithm for multimedia directional sensor networks," *International Journal of Communication Systems*, vol. 28, pp. 1598–1609, 2015.

[25] Y. Zhang, W. Luo, E. A. Mack, and R. Maciejewski, "Visualizing the impact of geographical variations on multivariate clustering," *Computer Graphics Forum*, vol. 35, pp. 101–110, 2016.

[26] X. Huang, Y. Zhao, C. Ma, J. Yang, X. Ye, and C. Zhang, "Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, pp. 160–169, 2016.

[27] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. Zhang, and J. Zhang, "Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, pp. 270–279, 2016.

[28] P. Rong, Z. Zheng, M.-P. Kwan, and Y. Qin, "Evaluation of the spatial equity of medical facilities based on improved potential model and map service API: A case study in Zhengzhou, China," *Applied Geography*, vol. 119, 2020.

Ke Zhang [M'09] received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), in 2010. He is currently an Associate Professor with the School of Computer Science and Engineering, UESTC, Chengdu, China. Ke Zhang has published more than 60 research articles in many journals and conferences. He is also the reviewer of many prestigious international conferences and journals, and he owns more than 14 national invention patents. His research interests include the internet of things, AI, big data, computer networks, and data fusion. His scientific research at UESTC started with designing an intelligent firewall system based on an Intel IXA network processor. Much of his focus has been on sensor networks, AI, and big data since 2006, participated in several research projects, such as the National Natural Science Foundation of China, the Sichuan Science and Technology Program and so on. He is also a Senior Member of the China Computer Federation (CCF) and the CCF's Member of the committee of experts on computer applications.

Hao Chen received the master degree from School of Communication and Information Engineering at University of Electronic Science and Technology of China. He is a software engineer working at Alibaba Group. His interest includes data visualization and data mining.

Hong-Ning Dai [SM'16] is currently with Department of Computing and Decision Sciences, Lingnan University, Hong Kong, China as an associate professor. He obtained the Ph.D. degree in Computer Science and Engineering from Department of Computer Science and Engineering at the Chinese University of Hong Kong. His current research interests include Internet of Things, big data, and blockchain technology. He has served as associate editors/editors for IEEE Transactions on Industrial Informatics, IEEE Systems Journal, IEEE Access, Ad Hoc Networks, and Connection Science.

Hongbo Liu is a Professor in University of Electronic Science and Technology of China since 2020. Before that, he was an Associate Professor in Indiana University Purdue University Indianapolis, USA. He obtained his PhD degree in electrical engineering from the Stevens Institute of Technology in 2013. His research interests include wireless sensing & mobile computing and cyber security & privacy. He is the recipient of the Best Paper Award from ACM MobiCom 2011 & IEEE CNS 2018, and Best Paper Runner-up Award from IEEE CNS 2013.

Zhongrui Lin received the B.E. degree in Intelligent Science and Technology from Nanjing University of Science and Technology (NJUST), in 2020. And now he is pursing the MA.Eng degree in Electronic Information at University of Electronic Science and Technology of China (UESTC). In the undergraduate stage, he studied the searchable encryption technology and separately published a paper on the 18th International Conference on Cryptology and Network Security (CANS2019) and the 2nd International Conference on Machine Learning for Cyber Security (ML4CS2019). He is now working on digital twin, smart city and machine learning.