



A Data Driven Approach to Measure Evolution Trends of City Information Modeling



Guangdong Wu¹, Handong Tang¹, Yichuan Deng^{2,3}, Hengqin Wu⁴, Chaoran Lin^{5*}

¹ School of Public Policy and Administration, Chongqing University, 400044 Chongqing, China

² School of Civil and Transportation Engineering, South China University of Technology, 510641 Guangzhou, China

³ State Key Laboratory of Subtropical Building Science, South China University of Technology, 510641 Guangzhou, China

⁴ College of Civil and Transportation Engineering, Shenzhen University, 518060 Shenzhen, China

⁵ School of Economics & Management, Harbin Engineering University, 150001 Harbin, China

* Correspondence: Chaoran Lin (linchaoran@hrbeu.edu.cn)

Received: 06-18-2022

Revised: 07-22-2022

Accepted: 08-09-2022

Citation: G. W. Wu, H. D. Tang, Y. C. Deng, H. Q. Wu, and C. R. Lin, "A data driven approach to measure evolution trends of city information modeling," *J. Urban Dev. Manag.*, vol. 1, no. 1, pp. 2-16, 2022. <https://doi.org/10.56578/judm010102>.



© 2022 by the authors. Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: This work aims to reveal the current status of the city information modeling (CIM) from massive patent data, using the latent Dirichlet allocation (LDA) model, and quantify the evolution trends of future topics by the Hidden Markov Model (HMM). The results show that the CIM technologies can be divided into 17 topics. At the present stage, the technologies related to the Internet of things (IoT), big data and data management are the focus of the research and development (R&D) of CIM patents. Compared with the software technology, further development is needed for the hardware technology supporting CIM, particularly in terms of information acquisition (cameras and sensors), storage, and information transmitters. This study deepens the understanding of the CIM-related technical categories, and clarifies the direction of the development and evolution of CIM technology, providing a strong support to decision-makers in urban management.

Keywords: Smart city; City information modeling (CIM); Latent Dirichlet allocation (LDA); Hidden Markov model (HMM); Patented technology

1. Introduction

Cities now account for 54% of the world's population, consume 80% of natural resources, and emit 80% of greenhouse gases [1-3]. Rising urbanization promotes economic and social development while posing significant challenges to long-term urban development [4]. The topic of smart cities, which are distinguished by smart governance and smart growth, has gradually piqued the interest of researchers seeking to promote sustainable urban development [5].

In the last decade, the centralized outbreak of networks, big data, artificial intelligence, modeling, and other technologies, particularly the development of sensors and low-power wide-area network technology, enabled the accurate and real-time representation of the physical world's dynamics in digital format. City Information Modeling (CIM) emerged in response to the concept of digital twins [6, 7].

A digital twin is a virtual entity and subsystem that characterizes a physical device in virtual space using the data from the physical device. CIM is a specific application of digital twin technology in urban management that lays the groundwork for smart city construction [8]. CIM enables data granularity down to a single module within a city building, transforming the traditional static digital city into a perceptible, dynamic, interactive, and intelligent real-world city. As a result, CIM provides critical data support for comprehensive urban management and fine governance [9], becoming an essential cornerstone for smart city operations.

CIM research typically focuses on CIM application scenarios in urban management [8, 10, 11], with few studies of CIM-related technology development and even fewer studies of patent literature research on CIM technology development. Using a scientific analytical framework to efficiently analyze patent documents, gaining technical

knowledge, and connecting problems with potential solutions all play important roles in encouraging more effective innovation in this field [12].

The ultimate goal of smart cities is to adhere to sustainable urban development principles and accomplish effective urban management (e.g., urban planning, infrastructure, transportation, energy, services, education, health, and public safety) while meeting public needs [3]. CIM, as a smart city management digital twin system, leverages information and communication technology to make critical components of urban infrastructure and services more interactive, accessible, and effective [13]. As a result, in order to support the development of smart cities, it is vital to understand the technological development trends in CIM.

Patent documents are a valuable source of technical information since their contents are accurate, thorough, and cutting-edge [14]. Patent data is increasingly being used by academics to investigate technological development trends. Pellicer et al. [15] discovered, for example, that basic data gathering, processing, and transmission technologies are required for smart city applications. Furthermore, Wang et al. [16] discovered that network communication and administration technologies are an important development path in the digital twin area by researching digital twin patent data.

However, due to the limits of research methods, previous studies failed to investigate the distribution of topics and the link between topics from a complete standpoint. The building of CIM by municipal managers, in particular, is not restricted to offering services autonomously. Rather, it necessitates the deployment of a comprehensive infrastructure for urban data collecting, transfer, storage, and analysis in order to provide public services [17]. In this context, this study aims to close the gap by throwing light on the technical scope of CIM and investigating its development trend. This study specifically wants answers to the following research questions:

- I. What is the most important technology in the field of CIM?
- II. What are the most essential CIM-related technologies?
- III. What is the goal and direction of future CIM-related technology development?

The work classifies patent literature data by using latent Dirichlet allocation (LDA) and the Hidden Markov Model (HMM) to discover the most relevant CIM technologies. From the standpoint of technology change, the changes in content and co-occurrence of technology topics are explored, technology trends are forecasted, and visual displays are developed. CIM-related businesses can shorten the research and development (R&D) cycle, reduce R&D expenses, comprehend the present technological environment, and gain insight into market trends by researching the evolution of CIM technology themes [18]. Additionally, it assists in defining the course of CIM technology development and evolution and directs industries to occupy the next technological highlands [19].

2. Literature Review

2.1 Smart Cities

The "smart growth" movement of the 1990s is where the idea of a "smart city" first emerged [5]. Since then, it has expanded to include nearly every type of cutting-edge technological application for urban planning, construction, operation, and management, including traffic improvement [20, 21], environmental sustainability [22], and urban governance [1, 23]. Despite the idea of smart cities becoming more and more popular, no consensus exists on what constitutes a smart city. According to studies by Lara et al. [24] and Mora et al. [25], the most widely used definition of a smart city is a community that systematically promotes the overall well-being of all its members and has sufficient flexibility to adapt and become a better place to live, work, and play in a sustainable manner.

2.2 CIM

CIM is a sophisticated synthesis of urban information sources and three-dimensional spatial modeling [26, 27]. In a limited sense, it consists of large-scale GIS and BIM data from the development of smart cities [27]. Based on the integration of BIM and GIS technologies, CIM enables data granularity precise to a single module inside the city building, creating an intelligent city and providing crucial data support for thorough urban administration and excellent governance [9].

Unified data standards, urban information models, urban operation data, common supporting platforms, and digital twin applications are just a few of the numerous components that make up a smart city. CIM serves as the key connection in this process by creating a digital twin of a real city model. Digital twin construction emerges as a significant driving force to modernize the capacity of urban governance. It relies on the three-dimensional digital urban backplane of the CIM platform and is highly integrated with real-time perception, simulation, deep learning, and other information technology to carry out omni-dimensional and multi-dimensional smart city application.

To achieve cross-system application integration and cross-departmental information sharing and support the decision-making analysis of smart cities, CIM's extensibility promotes accessing the information resources of many urban public systems (such as population, housing, household water, electricity and gas information, security police data, traffic information, tourism resource information, and public health care).

2.3 Topic Extraction of Patent Documents

Traditional techniques evaluate exemplary patents and extract technical topics from patent documents using expert knowledge [18]. The lack of expert resources, the challenge of determining patent representativeness, and the inability to study a large number of patents limit the usefulness of this strategy. Albino et al. [14] used the categorization properties of patents, such as the International Patent Classification number, as their technical topics to examine the evolution features of a particular field, trying to avoid relying on expert experience.

The limited types of patent features, however, compromise the precision of evolution trend analysis. Researchers used the patent co-occurrence network and citation relationship as a strategy to increase the accuracy of patent topics extraction [28]. The timeliness of topic evolution trend analysis cannot be guaranteed by this method due to the delays it produces. Therefore, when mining technical topics from patents and other scientific or technological documents, researchers use the Subject-Action-Object (SAO) structures in semantic similarity recognition, topic modeling, or topic clustering to account for the diversity of technical topics and the timeliness of the analysis [29-31].

A flexible probabilistic model called LDA was created specifically for topic modeling [32]. In order to represent the process of creating documents, this model uses probabilistic latent semantic analysis to introduce Dirichlet prior distribution and maximize word co-occurrence probability to look for word clustering. Consequently, it clusters documents and effectively extracts hidden topics. However, LDA's dimensionality reduction effect and recognition rate are compromised by its inability to keep the local structural information adequately.

To ensure effective LDA performance, an ideal number of topics must be chosen using a scientific manner. According to a 2012 proposal by Blei [33], the number of themes should be gauged by how perplexing they are. The language probability model's performance can be assessed and its parameters can be improved using perplexity. Perplexity determines the geometric mean of the sentence similarity in the literature set and estimates the information entropy of the probability distribution based on the information theory. Therefore, this work uses perplexity to evaluate and determine the optimal number of topics in the sample literature.

2.4 Topic Evolution of Patent Documents

Grey prediction [34] and time series analysis are common methods used by researchers to predict the topics' evolution tendency [35]. However, these analysis techniques typically disregard the random characteristics of the innovation process, despite the fact that randomness is a crucial element of innovation [36, 37]. Ignoring randomness in the quantitative forecast of future technological trends causes an overestimation of the endurance of current technical themes and an underestimation of the exponential emergence of new technologies.

The topic evolution may be influenced by two different processes. The first is the motivation that academics gain from reading about past research breakthroughs and the introduction of fresh concepts during literary change. But because there aren't many records, this process is thought to be a secret sequence that can't be seen. Second, motivated by the first step, researchers efficiently document the research findings in scientific literature, producing observable sequences. The latter process serves as the former's micro-foundation, and the former serves as the latter's macro-performance. Therefore, topic evolution can be seen as the superposition of the two processes.

Baum and Petrie's HMM, a probability and statistical model, was developed in 1966 as a method to capture such superposition [38]. The concept was initially employed in language processing before becoming well-known and being adopted in other disciplines.

3. Methodology

3.1 Data Sources

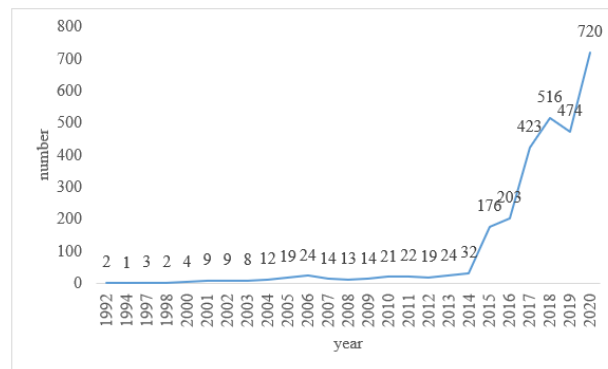


Figure 1. Number of CIM patents from 1990 to 2020

The frequency approach is employed in this study to choose keywords. To create a trustworthy keyword search list, 48 CIM-related articles from 2018 to 2020 were first chosen, and their topics, titles, abstracts, and keywords were examined. Second, additional data was used to extract keyword synonyms and assess the significance of the keywords (using sources like Wikipedia). Four keywords were selected after screening the frequency of CIM-related terms.

On February 20, 2021, "Patentscope" was searched the World Intellectual Property Organization (WIPO) patent database for patents relating to CIM using the following search terms: TS="city information model," "urban information model," "urban digital twins," and "digital city." As a result, 2764 CIM-related patents published between April 1992 and February 2021 were found. According to the period of application, Figure 1 displays the number of CIM patents that have been registered. As depicted in the image, CIM-related patents began to grow rapidly around 2015 and have continued to do so ever since.

3.2 Method Design

LDA is used in this study to model the data from the patent literature. In order to avoid the inefficiency and errors associated with human labeling, the model is first trained to create topics. The perplexity and transition matrices between topics in the evolution of CIM topic evolution are then computed using the state transition matrix in HMM and the probability distribution of the initial state. Finally, it is established how CIM topics have evolved over time and what the future evolution trend will be (Figure 2).

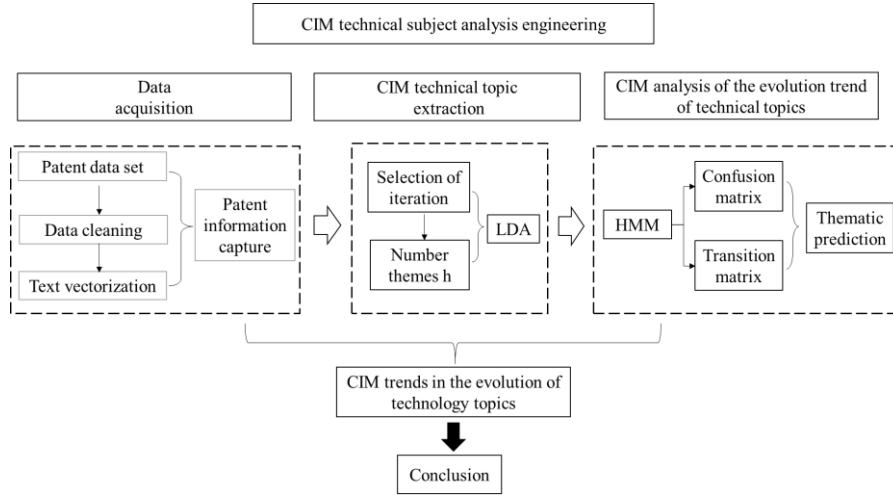


Figure 2. Workflow of the research

3.2.1 Document topic extraction module

The LDA algorithm's assumptions and requirements are followed in this study. We therefore suppose that each topic follows the hyper parametric Dirichlet prior distribution:

$$Dir(\theta_d|\alpha) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{dk}^{\alpha_k - 1} \quad (1)$$

where, θ_{dk} is the distribution of scientific documents d on topic k ; α is the distribution of subject words.

The topic term distribution $\phi_k \sim Dir(\beta)$ is generated for each topic k , and the topic term distribution $\theta_d \sim Dir(\alpha)$ is generated for each patent document d . Furthermore, the topic $Z_{dn} \sim Multinomial(\phi_{Z_{dn}})$ is derived for the n -th term in each document. Therefore, the LDA likelihood model can be established as:

$$p(W|\alpha, \beta) = \prod_{d=1}^D \int p(\theta_d|\alpha) \prod_{n=1}^{N_d} \sum_{Z_d} p(Z_{dn}|\theta_d) p(w_{dn}|\phi_{Z_{dn}}) d\theta_d \quad (2)$$

Perplexity can be calculated as:

$$Perplexity(D) = \exp\left(-\frac{\sum_{d=1}^M \log_D p(w_d)}{\sum_{d=1}^M N_d}\right) \quad (3)$$

$$p(w_d) = \sum_d \prod_{n=1}^T \sum_{j=1}^T p(w_j | z_j = j) \cdot p(z_j = j | w_d) \cdot p(d) \quad (4)$$

where, D is the test set in the corpus; M is the number of documents; N_d is the number of words in document d ; $p(w_d)$ is the probability of the occurrence of w_d .

To avoid over fitting, the number of topics and perplexity must be selected carefully. The number of topics associated with the lowest perplexity is considered as the optimal value in LDA model training.

Following Heinrich's parameter estimation, this work sets $\alpha = 50/k$ and $\beta = 0.1$. Furthermore, the Gibbs Sampling is used to derive the topic set $K = \{k_1, \dots, k_h\}$, and the topic attribution set of each document $D_k = \{j_1, \dots, j_n\}$.

3.2.2 Trend analysis module

A complete HMM model can be described as a tuple $\gamma = (S, \pi, A, B, O)$. The random transition sequence of hidden states can be expressed as:

$$S = \{s_1, \dots, s_h\} \quad (5)$$

Let $Q = \{q_1, \dots, q_t\}$ denote the randomly generated hidden state sequence, where $q_t \in S$, and t is the number of topics in the hidden state. The change of the hidden state represents the topic state change. Thus, Q represents the set of all possible topic states.

Then, the initial state of the system follows the probability distribution below:

$$\pi = \{\pi_i, 1 \leq i \leq N\} \quad (6)$$

where, π_i is the occurrence probability of state S_i . The probability distribution of transitions of the research topic from state S_i to S_j can be described as follows:

$$A = \{\alpha_{ij}\} \quad (7)$$

where, $\alpha_{ij} = P\{(q_{t+1} = S_j | q_t = S_i)\}$, $1 \leq i, j \leq N$, and satisfies $\alpha_{ij} \geq 0$, $\sum_{j=1}^N \alpha_{ij} = 1$. For state S_i , the probability distribution of the observed variables can be expressed as:

$$B = \{b_i(v)\} = \{f(Q_t = v | q_t = S_i)\} \quad (8)$$

where, Q_t is the t -th observed variable. Thus, the observation sequence can be expressed as:

$$O = \{O_1, \dots, O_t\} \quad (9)$$

where, the observed state at time t can be described as a vector $O_t = (\alpha(t), \beta(t))$, with $\alpha(t)$ and $\beta(t)$ being the inflow and outflow frequencies at time t , respectively. Therefore, the vector sequence formed by the inflow and outflow of all observed samples is the observation state O . Figure 3 shows the relationship between hidden state transition sequence and observation sequence.

The initial training value of the model is set as $O = Q = \{pt_1, \dots, pt_2\}$. Then, the Baum-Welch algorithm is used to estimate the model parameters, yielding a single optimal state sequence. The structure of the research topic after k years is obtained by $\hat{O}_{tk} = \sum_{j=1}^N A^K(i, j) E(b_j(v))$.

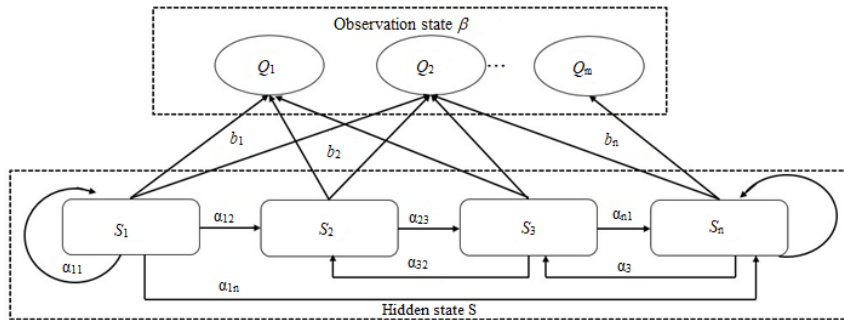


Figure 3. Relationship between hidden state transition sequence and observation sequence

4. Results

4.1 Topic Extraction of CIM Patent Documents

Table 1. Topics and their corresponding keywords

No.	KW1	KW2	KW3	KW4	KW5	ID	Category
1	Module	Control	Communication	Signal	Wireless	Communication control module	Network communication technology
2	Information	Management	Monitor	Terminal	Server	Information terminal	Network communication technology
3	Datum	Model	City	Information	Urban	Datum model	Heterogeneous application integration technology
4	Network	Model	Information	Prediction	Invention	Network model	Heterogeneous application integration technology
5	Information	Model	Accord	Feature	Target	Target recognition	IOT technology
6	Power	Electric	Control	Circuit	Energy	Circuit control	Network communication technology
7	Layer	Storage	Mechanism	Plurality	Computer	Plurality storage	Big data and data management technology
8	Card	Camera	Ring	Utility	Sign	Camera signal	IOT technology
9	Plate	End	Fix	Rod	Top	Plate fixed	IOT technology
10	Case	Computing	Electronic	Equip	Utility	Utility Computing	Cloud computing technology
11	Display	Screen	Board	Bus	Information	screen board	IOT technology
12	Vehicle	Traffic	City	Parking	Road	Vehicle road	IOT technology
13	Body	Machine	Box	Equip	Drive	Body equip	IOT technology
14	Sensor	Temperature	Utility	Body	Garbage	Temperature sensor	IOT technology
15	Computer	Computing	Automatic	Message	Drive	Autonomic Computing	Cloud computing technology
16	Lamp	Street	Light	City	Part	Lamp part	IOT technology
17	Water	Pipe	Gas	Detection	Pipeline	Pipeline detection	IOT technology

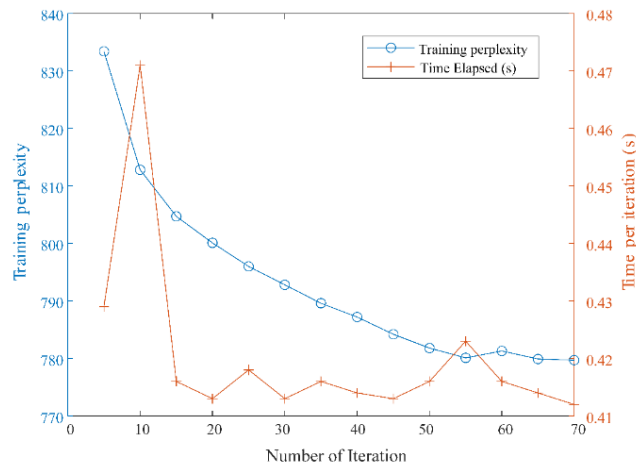


Figure 4. Effect of number of iterations on machine learning

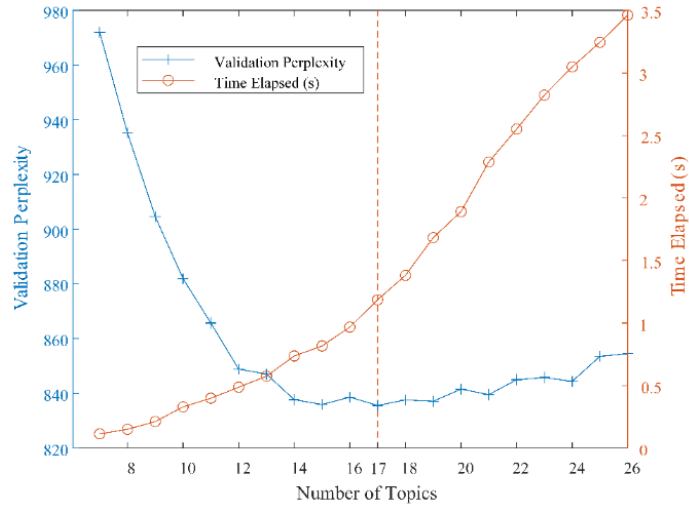


Figure 5. Perplexity scores for different number of topics

The learning effect of the LDA model, a machine learning technique, is highly correlated with the number of iterations. When there are more than 70 iterations in this study, the new iteration's contribution to the Log-Likelihood increase is almost zero. In order to reduce the cost of operation time, this study sets the number of iterations at 70. (Figure 4). The number of topics was also adjusted from six to seventy, and perplexity scores were computed. There is a minimum perplexity of 835.8 when there are 17 topics (Figure 5). Consequently, 17 topics were chosen.

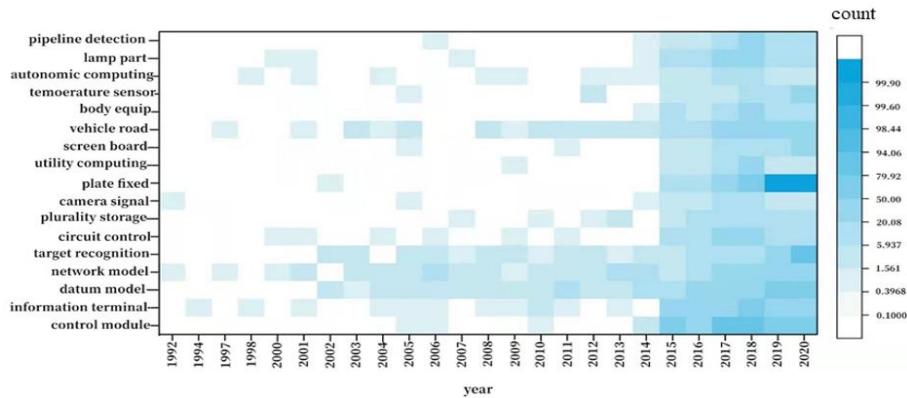
The LDA library (genism) in Python is used in this study to calculate the topic information and provide keywords for each topic. For each topic, 40 keywords are explicitly extracted. The top five keywords with the highest likelihood are chosen to represent the topic after the words are sorted based on the assessed probability. Each topic is labeled and categorized in accordance with the sort of technology it represents to ease learning and ease referencing (Table 1).

4.2 Evolution Path of CIM Patent Topics

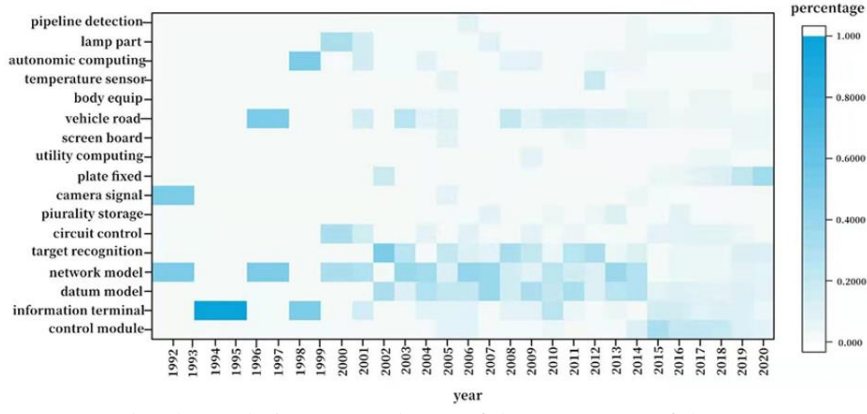
4.2.1 Temporal analysis

A heat map of the topic number and each topic's individual gravity trend is created based on the results that were shown (Figure 6). In Figure 6b, the annual change trend and initial occurrence time of the topics show how the topics are continually subdivided as the research progresses. The first CIM application dates back to the 1994 Amsterdam experiment De Digitale Stad [39]. At this point, network communication data utilization and urban data collection have been investigated, thus technological topics like network model, information terminal, and camera signal are developing [40].

With the advancement of network communication technologies from wired to wireless and mobile wireless networks, mobile information communication capabilities have been significantly improved. The widespread adoption of 2.5G, 3G, 4G, and 5G technologies, which provide network infrastructure for the quick development of new technologies, has had a significant impact on these advancements. Circuit control, cloud computing, and autonomous computing are topics in network communication.



(a) The evolutionary trend map of the number of themes



(b) The evolutionary trend map of the percentage of themes

Figure 6. Evolution charts of the number and proportion of topics

Additionally, the IOT technological topic known as lamp part has been created [8, 41]. The IOT has enabled the CIM information interaction mode, which primarily relies on human-computer interaction, to progress into the ubiquitous computing stage of multi-source real-time information acquisition and intelligent control [42]. As a result, further discussions evolved on the themes of target recognition, temperature sensors, and communication control module.

Parallel to this, the growth of the IOT and its applications has created a number of issues, including the administration, processing, fusion, integration, and mining analysis of huge data from several sources, giving rise to the topic of plurality storage [43-45]. The storage, administration, processing, sharing, integration, and application of multi-source enormous data present a considerable demand on computer resources with the development of a digital city or smart city. This problem is solved by cloud computing technology, which also gave rise to key topics like utility computing and autonomous computing.

4.2.2 Evolution path of CIM topics

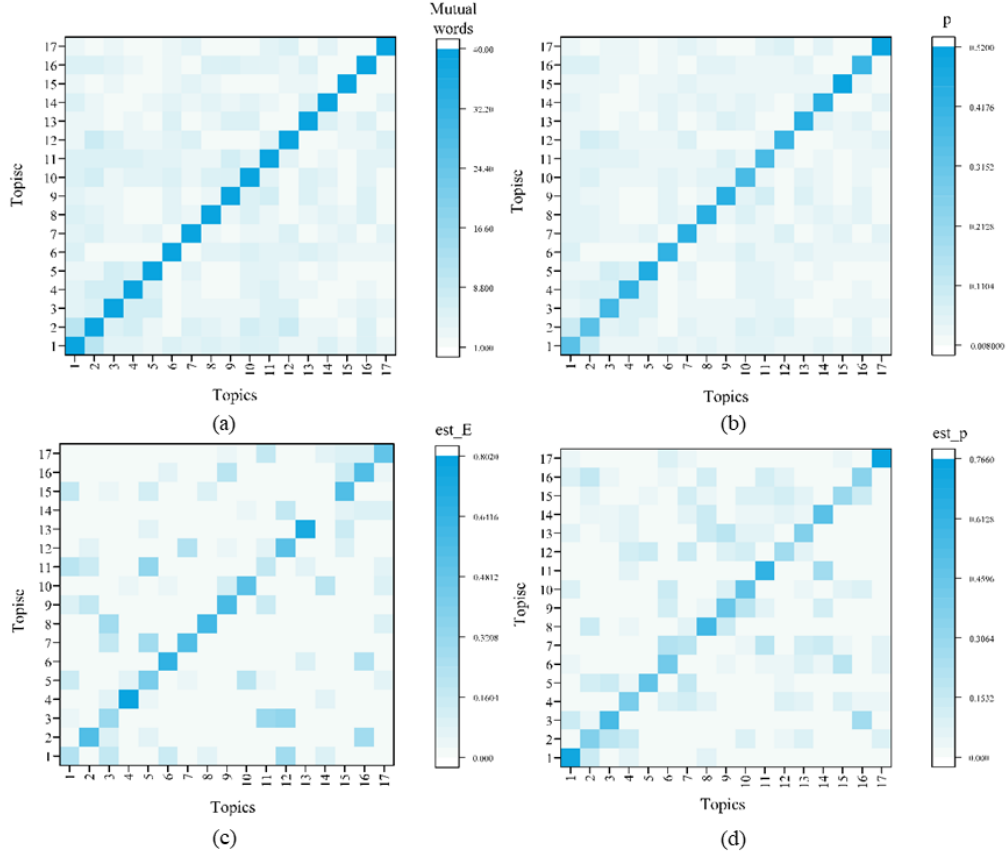


Figure 7. Heat map of word co-occurrence symmetry matrix

The likelihood of perplexity or transition between topics increases with the similarity of the topics. This paper bases its method of topic word co-occurrence analysis on this idea. We measure and use as a proxy for the degree of similarity across topics the co-occurrences of the first 40 topic words in 17 topics. To this end, the 17 word co-occurrence symmetry matrix was built, and the related heat map is shown in Figure 7. The diagonals are highlighted the most since each topic occurs with itself most frequently, whereas the other parts have varied hues because other topics occur less frequently. The likelihood that a hidden state would be detected as an observable state is displayed in the perplexity matrix. The threshold of transition between patent topics in the evolution process is measured using this probability. The matrix also indicates the likelihood and direction of a topic change at the same time. The perplexity matrix's dark squares correspond to the topics that are subject to change (transition) throughout time. It is apparent from Figure 7 that most CIM topics are straightforward, with a high degree of independence of research content and a high threshold for state transition. In other words, there is a high and clear barrier between most topics.

Topics having a perplexity probability of more than 10% are shown as a perplexity network diagrams to further the analysis (Figure 8). According to Figure 8, CIM R&D topics can go more readily from multi-component topics (e.g., drive equipment and information terminal)) to technology topics including multi-scene applications (such as screen board, temperature sensor, and camera signal). Additionally, a number of technology topics serve as crucial links in the development and transformation of other technology topics, serving as significant bridges and intermediaries in the growth of technology. The information terminal, for instance, serves as a node between the control module and the target recognition system and the vehicle road. Information terminal is a vital node in this transition because control module and target recognition are subdivisions of information terminal application, and vehicle road is the integration of information terminal application. A technical transition from target detection and automated drive to vehicle road is also provided by plurality storage. Important machine learning applications include target recognition and automated driving, and plurality storage offers a technical assurance for machine learning. Camera signal, vehicle road, and automatic driving are further topics with comparable functions. The ambiguity feature illuminates the prospective R&D development course of the CIM technology topic, making it simple for the micro R&D department to select the technology path scientifically while planning the scientific research objective.

The chance of switching between topics in the macrohistorical R&D process is represented by the transition matrix. The diagonal in Figure 7d, however, is highlighted, indicating that the majority of topics continue to exhibit consistency in their research trends. The characteristics of the transition in R&D trend among topics vary when seen from the perspective of a single topic. In this study, a relationship feature diagram is plotted using the topics that have a transition probability of more than 15% (Figure 9).

Datum modeling, target identification, and screen boards are some of the topics with a great ability to retain their R&D direction. The topic with the largest proportion of transition loss throughout evolution is "screen board," which mainly flows to target recognition, control module, information terminal, and body equipment. These topics have consistently maintained a high degree of co-occurrence and are hot technology nodes in the CIM-related R&D. This behavior is brought on by the ongoing diversification of wireless terminal products, the escalating anti-interference standards for circuit boards, the steady decline in R&D for simple-use screen boards, and the rise in R&D and use of specialized equipment.

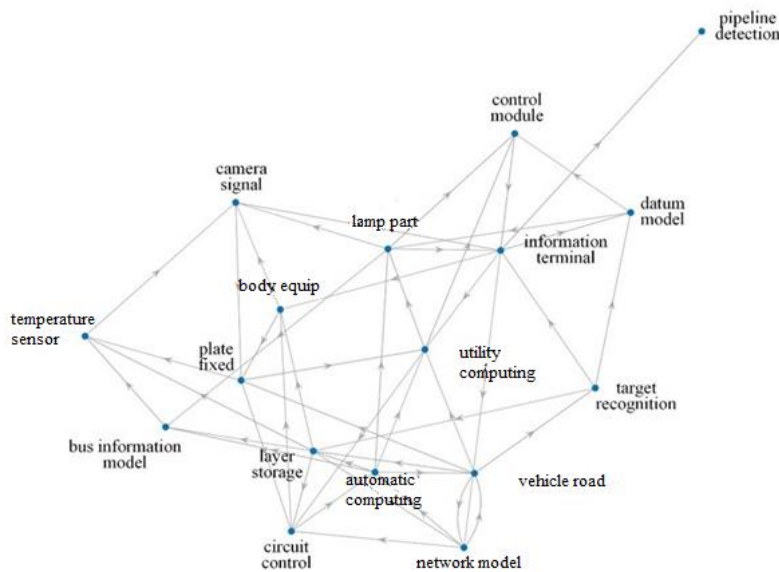


Figure 8. Perplexity relationship between topics

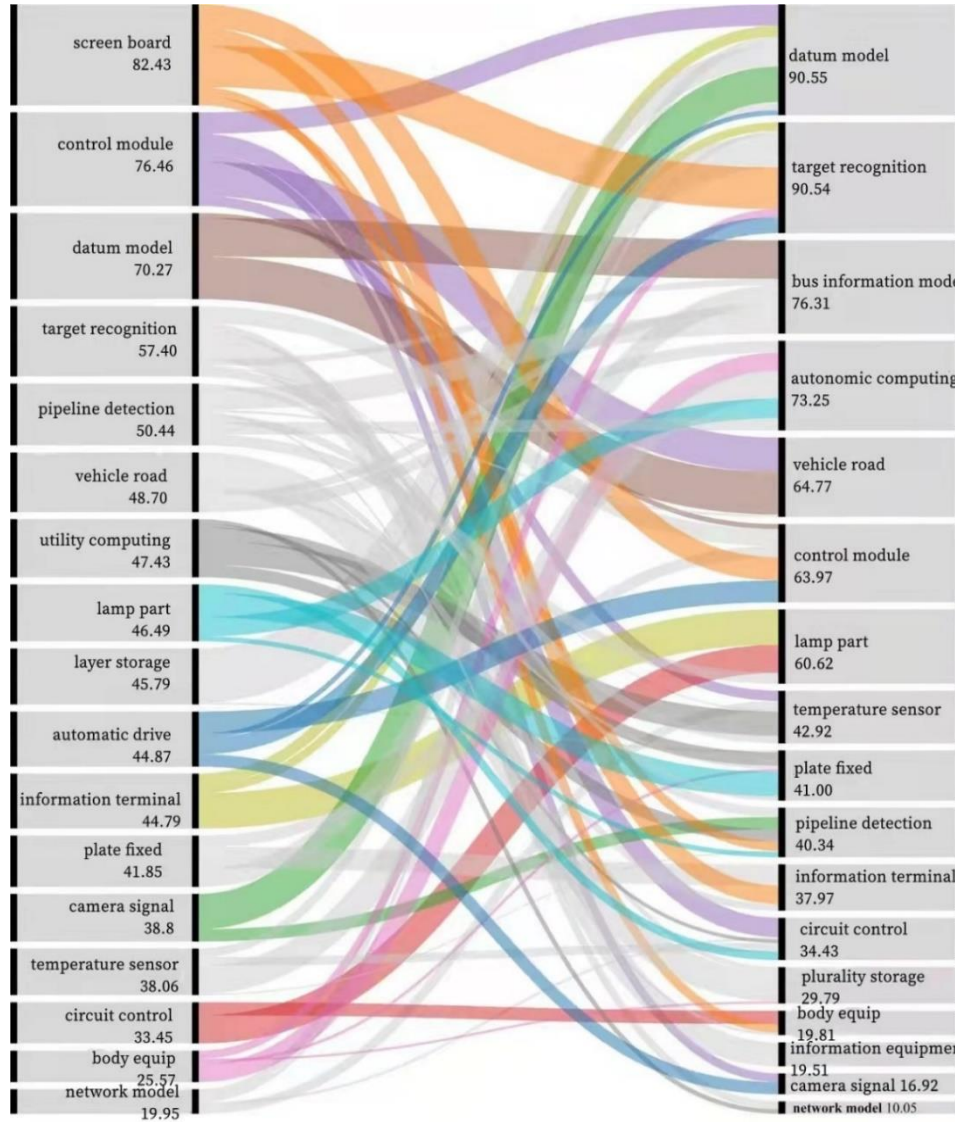


Figure 9. HMM estimation of transition probabilities between topics

Since the current CIM platform has not yet developed a unified spatiotemporal data underlying framework compatible with heterogeneous information systems, the "datum model" has seen the largest increase in popularity, and its R&D strength mainly comes from "camera signal," "information terminal," and "plurality storage." There are many different standard formats for data collecting or design modeling software, and there exist hurdles to data accommodation between formats. Multi-source data includes vector data, raster data, model data, point cloud data, and other data from several sources. The CIM role is constrained if the interdepartmental business system's data format is uncoordinated, the data authority is unclear, and the data docking mechanism is flawed. As a result, the study of the datum model is always expanding. Additionally important is the rise in the "heat" of target recognition. Target identification, a hot topic in the realm of deep learning, is a crucial way of data collecting and human-computer interaction in the CIM application. The foundational technology of an intelligent city is the road target identification algorithm, which employs deep learning to assess the condition of the road. Finding the target's position and size inside an image, as well as detecting particular target classes, are two of the functions of target recognition. Intelligent security, intelligent medical care, unmanned retail, intelligent hardware, and intelligent robotics are some of the related fields.

4.2.3 Prediction and analysis of CIM topics

The 2021 literature has not yet been released in its whole as of the conclusion of this study. Therefore, this study uses the patents released up until the end of 2020 as the training set to anticipate the CIM evolution trends, and it adds the predicted perplexity matrix and transition matrix to the Matlab HMM module. The outcomes of the HMM prediction of the evolution of the CIM topics from 2018 to 2026 are as a consequence obtained, as depicted in Figure 10.

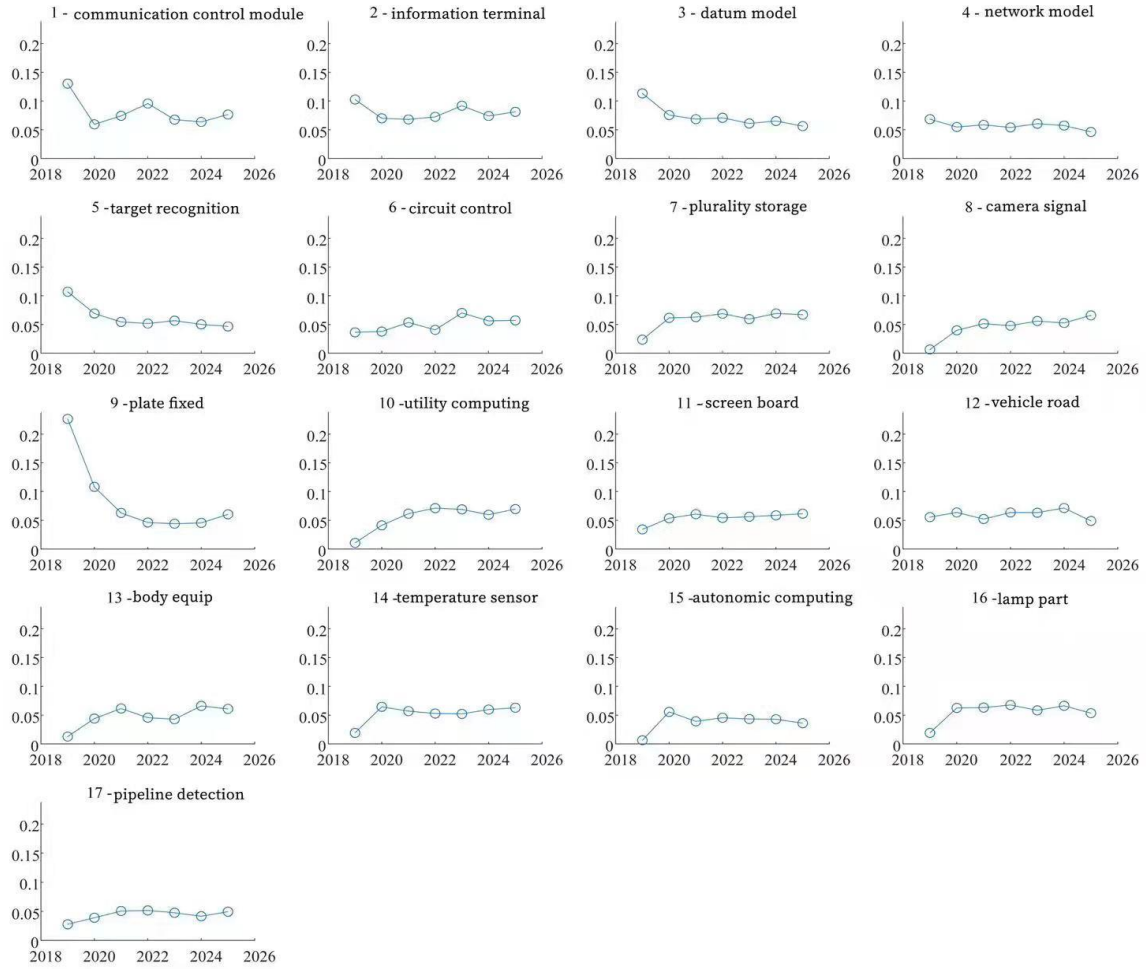


Figure 10. Predicted trends of CIM topics from 2019 to 2026

The hidden Markov predictions reveal a sharp rise in the proportion of topics related to temperature sensors, camera signals, electrical equipment, and body equipment. Front-end data collection is necessary for the development of CIM. Real-time data acquisition is encouraged by effective front-end data collection [8]. Artificial intelligence and machine learning technologies provide a better understanding of how cities evolve and cope with drawbacks as sensors, computing cores, and more effective electronic communication systems get installed in urban infrastructure. The Internet of Everything is made possible by the CIM platform [46, 47], and the growth of electrical and body equipment is in line with these developments. Radio frequency identification, infrared sensors, global positioning systems, laser scanners, and other information sensing devices are becoming increasingly embedded in power grids, railways, bridges, tunnels, roads, buildings, water supply systems, dams, oil and gas pipelines, and other objects, thanks to the advancement of front-end data collection systems. Micro-platform integration now replaces fixed technology (plate fixed) requirements for particular platform equipment.

Target recognition is an image recognition technology based on machine learning. The advancement of this technology necessitates increased quantity and quality of visual data as well as the development of more effective algorithms [48]. Target recognition hence entered the period of technology accumulation, after quick early development, when the pre-technology such as camera signal failed to progress considerably. As a result, the degree of R&D has somewhat decreased, calling for an upgrade and optimization of the relevant technology (such as machine learning supporting technology). Similar to this, it became challenging to create technical innovation without significant advances in hardware technology, only based on the early development of technologies related to network and datum models. As a result, related technology development is very sluggish.

5. Discussion

5.1 CIM Related Technology

The examination of the current patent topics in this work leads to the identification of technical CIM-related topics. Data collection, data transmission, data storage, and data application are the four main technological

domains that the technologies associated with CIM and Smart City fall under [15]. CIM views the city as a whole, or as an organic system capable of intelligent and coordinated operation, from a technological perspective. Hence, the city is transformed into an ever-more-powerful digital system. Through the communication control module and the circuit control (nervous system), sensor signals like camera signal and temperature (sensory system) are embedded into the ubiquitous information terminal and datum model (brain), directing automatic drive, pipeline detection, and other data terminals. CIM, which is an essential component of managing smart cities, is the digital twin of urban entities on digital equipment [8]. However, because management technology is not included in the CIM construction as a concept of digital information technology, it is unable to completely actualize smart management even though it can assist urban intelligence management [3]. Along with digital information technology, management technologies such as policy-driven management [49] and social restrictions [50] should be taken into account when managing smart cities.

From a technological standpoint, the research of CIM now concentrates on IOT-related technologies (such as camera signals, electrical equipment, body equipment, and temperature sensors), big data, and data management technology (plurality storage). Mobile communication has greatly advanced with the adoption of 4G and 5G technologies, supplying the network infrastructure for the quick development of the IOT [42]. IOT technology advancements have resulted in an increase in data volume and data types in cities. This enhances the usefulness of and active development in associated data scheduling and management technologies (such as effective indexing, databases, and distributed storage).

5.2 Development Trends and Focus of CIM Technology

Two tendencies in the advancement of CIM technology were identified by the analysis of the existing patent literature. Namely, software technology progress tends to slack off while hardware technology advances steadily. These trends do not necessarily reflect the relative importance of hardware and software technology, but rather the inadequacies in the current hardware that prevent it from supporting the significant advancement of software technology. For some target recognition applications, high-resolution image technology is necessary. For instance, in autopilot systems, target recognition is crucial, making lidar and camera necessary. The topic transition diagram also demonstrated the flow of target reorganization technology research and development to other technology sectors, including video signal and communication control module research. This highlighted the necessity for additional study on automated driving technology. As a result, municipal administrators must raise their spending on hardware R&D to further the development and promotion of CIM. To encourage the creation of the related follow-up applications, the hardware technology must continuously advance.

6. Conclusions

City managers must transition from traditional urban administration to intelligent city management using high-tech tools as a result of the sharp increase in urban population. The growing demand for intelligent city management among city managers has accelerated the technological advancement of related services and goods, with the development of the associated CIM technology being essential to the creation of the smart city. Therefore, it is urgently needed to reassess CIM technology development. The LDA topic model and HMM are used in this study to offer a fresh approach to mining and forecasting the topics of scientific publications. The technique allows for speedy mining of hidden topic information, collecting the main subject nodes and topic evolution patterns, and enabling efficient and unsupervised (i.e., without expert input) clustering of literature. It also offers a novel approach to text analysis.

Our method divides the associated technologies in the field of CIM into 17 topics, with the R&D of CIM patented technology focused on plurality storage, camera signal, electronic equip, body equipment, and temperature sensor. The three most significant technological advancement directions in the future are front-end collecting technology, terminal application technology, and data management technology. The hardware supporting technology for CIM needs to be developed concurrently, in contrast to the software technology. The technology particularly covers information transmission equipment, storage equipment, and information collecting equipment (cameras, sensors, etc.). As these fields are still in the early stages of research, it is necessary to increase their capacity for innovation. Scientific research institutes from different nations will need to find a solution to the issue of how to conduct additional innovative research on these R&D topics in the future.

The limitations of this study, however, may serve as a guiding principle for future research. First and foremost, this article exclusively examined the patent information included in the WIPO Patentscope database. Additional patent information from various patent databases should be included in future studies. Second, this study classifies the related topics using the LDA model. Future studies will analyze CIM patent documents using various categorization techniques, such as the BERT model [51]. Finally, the disparities between various locations are not taken into consideration in the examination of technology development trends that has been presented. Future research will therefore compare CIM development trends in other nations.

Fundings

This study was supported by the National Natural Science Foundation of China (Grant No.: 71972018 and 72104064); the Fundamental Research Funds for the Central Universities (Grant No.: 2021CDJSKJC02 and 2022CDJSKPY13).

Data Availability

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. Arbolino, F. Carlucci, A. Cirà, G. Ioppolo, and T. Yigitcanlar, "Efficiency of the EU regulation on greenhouse gas emissions in Italy: The hierarchical cluster analysis approach," *Ecol. Indic.*, vol. 81, pp. 115-123, 2017. <https://doi.org/10.1016/j.ecolind.2017.05.053>.
- [2] R. Arbolino, F. Carlucci, A. Cirà, T. Yigitcanlar, and G. Ioppolo, "Mitigating regional disparities through microfinancing: An analysis of microcredit as a sustainability tool for territorial development in Italy," *Land. Use. Policy.*, vol. 70, pp. 281-288, 2018. <https://doi.org/10.1016/j.landusepol.2017.10.042>.
- [3] T. Yigitcanlar, M. Kamruzzaman, L. Buys, G. Ioppolo, J. Sabatini-Marques, E. M. da Costa, and J. J. Yun, "Understanding 'smart cities': Intertwining development drivers with desired outcomes in a multidimensional framework," *Cities.*, vol. 81, pp. 145-160, 2018. <https://doi.org/10.1016/j.cities.2018.04.003>.
- [4] S. Allwinkle and P. Cruickshank, "Creating smart-er cities: An overview," *J. Urban Technol.*, vol. 18, no. 2, pp. 1-16, 2011. <https://doi.org/10.1080/10630732.2011.601103>.
- [5] J. C. Griffith, "Smart governance for smart growth: The need for regional governments," *Ga. St. UL Rev.*, vol. 17, pp. 1019-1019, 2000.
- [6] R. P. Dameri and C. Benevolo, "Governing smart cities: an empirical analysis," *Soc. Sci. Comput. Rev.*, vol. 34, no. 6, pp. 693-707, 2016. <https://doi.org/10.1177/089443931561109>.
- [7] K. H. Soon and V. H. S. Khoo, "CityGML modelling for Singapore 3d national mapping," *Int. Arch. Photogramm. Remote Sens-Basel. Spat. Inform. Sci.*, vol. 42, pp. 37-42, 2017.
- [8] G. White, A. Zink, L. Codecá, and S. Clarke, "A digital twin smart city for citizen feedback," *Cities.*, vol. 110, Article ID: 103064, 2021. <https://doi.org/10.1016/j.cities.2020.103064>.
- [9] Y. Cai, H. Huang, K. Wang, C. Zhang, L. Fan, and F. Guo, "Selecting optimal combination of data channels for semantic segmentation in city information modelling (CIM)," *Remote Sens-Basel.*, vol. 13, no. 7, pp. 1367-1367, 2021. <https://doi.org/10.3390/rs13071367>.
- [10] T. Deng, K. Zhang, and Z. J. M. Shen, "A systematic review of a digital twin city: A new pattern of urban governance toward smart cities," *J. Manag. Sci. Eng.*, vol. 6, no. 2, pp. 125-134, 2021. <https://doi.org/10.1016/j.jmse.2021.03.003>.
- [11] A. El Saddik, "Digital twins: The convergence of multimedia technologies," *IEEE Multimedia.*, vol. 25, no. 2, pp. 87-92, 2018. <https://doi.org/10.1109/MMUL.2018.023121167>.
- [12] H. Wu, G. Q. Shen, X. Lin, M. Li, and C. Z. Li, "A transformer-based deep learning model for recognizing communication-oriented entities from patents of ICT in construction," *Automat. Constr.*, vol. 125, Article ID: 103608, 2021. <https://doi.org/10.1016/j.autcon.2021.103608>.
- [13] A. Alkandari, M. Alnasheet, and I. F. Alshaikhli, "Smart cities: A survey," *J. Adv. Comput. Sci. Tech Res.*, vol. 2, no. 2, pp. 79-90, 2012.
- [14] V. Albino, L. Ardito, R. M. Dangelico, and A. M. Petruzzelli, "Understanding the development trends of low-carbon energy technologies: A patent analysis," *Appl. Energ.*, vol. 135, pp. 836-854, 2014. <https://doi.org/10.1016/j.apenergy.2014.08.012>.
- [15] S. Pellicer, G. Santa, A. L. Bleda, R. Maestre, A. J. Jara, and A. G. Skarmeta, "A global perspective of smart cities: A survey," In 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Taichung, Taiwan, 03-05 July 2013, IEEE, pp. 439-444. <https://doi.org/10.1109/IMIS.2013.79>.
- [16] K. J. Wang, T. L. Lee, and Y. Hsu, "Revolution on digital twin technology-a patent research approach," *Int. J. Adv. Manu. Tech.*, vol. 107, no. 11, pp. 4687-4704, 2020. <https://doi.org/10.1007/s00170-020-05314-w>.
- [17] H. Samih, "Smart cities and internet of things," *J. Inform. Tech. Case. Appl. Res.*, vol. 21, no. 1, pp. 3-12, 2019.

- <https://doi.org/10.1080/15228053.2019.1587572>.
- [18] J. Wonglimpiyarat, "Technological change of the energy innovation system: From oil-based to bio-based energy," *Appl. Energ.*, vol. 87, no. 3, pp. 749-755, 2010. <https://doi.org/10.1016/j.apenergy.2009.08.043>.
 - [19] S. J. Liu and J. Shyu, "Strategic planning for technology development with patent analysis," *Int. J. Technol. Manage.*, vol. 13, no. 5-6, pp. 661-680, 1997. <https://doi.org/10.1504/IJTM.1997.001689>.
 - [20] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, "UAV-enabled intelligent transportation systems for the smart city: Applications and challenges," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 22-28, 2017. <https://doi.org/10.1109/MCOM.2017.1600238CM>.
 - [21] T. Yigitcanlar, L. Fabian, and E. Coiacetto, "Challenges to urban transport sustainability and smart transport in a tourist city: The Gold Coast, Australia," *Open Trans. J.*, vol. 2, pp. 29-46, 2008. <https://doi.org/10.2174/1874447800802010029>.
 - [22] M. D. Lytras, A. Visvizi, and A. Sarirete, "Clustering smart city services: Perceptions, expectations, responses," *Sustain.*, vol. 11, no. 6, pp. 1669-1669, 2019. <https://doi.org/10.3390/su11061669>.
 - [23] L. Parra, S. Sendra, J. Lloret, and I. Bosch, "Development of a conductivity sensor for monitoring groundwater resources to optimize water management in smart city environments," *Sensors-Basel.*, vol. 15, no. 9, pp. 20990-21015, 2015. <https://doi.org/10.3390/s150920990>.
 - [24] A. P. Lara, E. M. Da Costa, T. Z. Furlani, and T. Yigitcanla, "Smartness that matters: Towards a comprehensive and human-centred characterisation of smart cities," *J. Open Innovation. Tech. Market. Complexity.*, vol. 2, no. 2, pp. 8-8, 2016. <https://doi.org/10.1186/s40852-016-0034-z>.
 - [25] L. Mora, R. Bolici, and M. Deakin, "The first two decades of smart-city research: A bibliometric analysis," *J. Urban Technol.*, vol. 24, no. 1, pp. 3-27, 2017. <https://doi.org/10.1080/10630732.2017.1285123>.
 - [26] T. Stojanovski, "City Information Modelling (CIM) and Urban Design-Morphological Structure, Design Elements and Programming Classes in CIM," *eCAADe*, vol. 2018, pp. 507-516, 2018. <https://doi.org/10.52842/conf.ecaade.2018.1.507>.
 - [27] X. Xu, L. Ding, H. Luo, and L. Ma, "From building information modeling to city information modeling," *J. Inf. Technol. Constr.*, vol. 19, pp. 292-307, 2014.
 - [28] A. Magrí, F. Giovannini, R. Connan, G. Bridoux, and F. Béline, "Nutrient management from biogas digester effluents: a bibliometric-based analysis of publications and patents," *Int. J. Environ. Sci. Te.*, vol. 14, no. 8, pp. 1739-1756, 2017. <https://doi.org/10.1007/s13762-017-1293-3>.
 - [29] C. G. Figuerola, F. J. García Marco, and M. Pinto, "Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA," *Scientometrics.*, vol. 112, no. 3, pp. 1507-1535, 2017. <https://doi.org/10.1007/s11192-017-2432-9>.
 - [30] B. Hu, X. Dong, C. Zhang, T. D. Bowman, Y. Ding, S. Milojević, and V. Larivière, "A lead-lag analysis of the topic evolution patterns for preprints and publications," *J. Assoc. Inf. Sci. Tech.*, vol. 66, no. 12, pp. 2643-2656, 2015. <https://doi.org/10.1002/asi.23347>.
 - [31] H. Jiang, M. Qiang, and P. Lin, "Finding academic concerns of the Three Gorges Project based on a topic modeling approach," *Ecol. Indic.*, vol. 60, pp. 693-701, 2016. <https://doi.org/10.1016/j.ecolind.2015.08.007>.
 - [32] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993-1022, 2003.
 - [33] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, pp. 77-84, 2012. <https://doi.org/10.1145/2133806.2133826>.
 - [34] W. Li, "Application of grey prediction theory to forecast technology input within the Chinese High-Tech Industries," *Int J. Adv. Compu. Technol.*, vol. 4, no. 12, 2011. <https://doi.org/10.1109/ICACC.2011.6016373>.
 - [35] G. E. Heo, K. Y. Kang, M. Song, and J. H. Lee, "Analyzing the field of bioinformatics with the multi-faceted topic modeling technique," *BMC Bioinformatics.*, vol. 18, no. 7, pp. 45-57, 2017. <https://doi.org/10.1186/s12859-017-1640-x>.
 - [36] A. Hargadon and R. I. Sutton, "Technology brokering and innovation in a product development firm," *Admin. Sci. Quart.*, vol. 42, no. 4, pp. 716-749, 1997. <https://doi.org/10.2307/2393655>.
 - [37] M. Rogers and M. Rogers, *The definition and measurement of innovation*, Parkville, VIC: Melbourne Institute of Applied Economic and Social Research, 1998.
 - [38] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math Stat.*, vol. 37, no. 6, pp. 1554-1563, 1966.
 - [39] L. Francissen and K. Brants, "Virtually going places: Square hopping in Amsterdam's digital city," In *Cyberdemocracy*, R. Tsagarousianou, D. Tambini, and C. Bryan (Eds.), London: Routledge, pp. 18-40, 1997.
 - [40] J. Dykes, G. Andrienko, N. Andrienko, V. Paelke, and J. Schiewe, "Editorial—GeoVisualization and the digital city," *Comput. Environ. Urban.*, vol. 34, no. 6, pp. 443-451, 2010. <https://doi.org/10.1016/j.compenvurbsys.2010.09.001>.
 - [41] G. S. Yovanof and G. N. Hazapis, "An architectural framework and enabling wireless technologies for digital cities & intelligent urban environments," *Wireless Pers. Commun.*, vol. 49, no. 3, pp. 445-463, 2009. <https://doi.org/10.1007/s11277-009-9693-4>.

- [42] M. Batty, "Big data and the city," *Built Environ.*, vol. 42, no. 3, pp. 321-337, 2016. <https://doi.org/10.2148/benv.42.3.321>.
- [43] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *Natl. Sci. Rev.*, vol. 1, no. 2, pp. 293-314, 2014. <https://doi.org/10.1093/nsr/nwt032>.
- [44] C. Lim, K. J. Kim, and P. P. Maglio, "Smart cities with big data: Reference models, challenges, and considerations," *Cities.*, vol. 82, pp. 86-99, 2018. <https://doi.org/10.1016/j.cities.2018.04.011>.
- [45] A. A. Tole, "Big data challenges," *Database Sys. J.*, 4, no. 3, pp. 31-40, 2013.
- [46] S. E. Bibri, "The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability," *Sustain. Cities Soc.*, vol. 38, pp. 230-253, 2018. <https://doi.org/10.1016/j.scs.2017.12.034>.
- [47] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in Smart City initiatives: Some stylised facts," *Cities.*, vol. 38, pp. 25-36, 2014. <https://doi.org/10.1016/j.cities.2013.12.010>.
- [48] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities.*, vol. 89, pp. 80-91, 2019. <https://doi.org/10.1016/j.cities.2019.01.032>.
- [49] A. Caragliu and C. Del Bo, "Smartness and European urban performance: assessing the local impacts of smart urban attributes," *Innovation. Eur. J. Soc. Sci. Res.*, vol. 25, no. 2, pp. 97-113, 2012. <https://doi.org/10.1080/13511610.2012.660323>.
- [50] I. Beretta, "The social effects of eco-innovations in Italian smart cities," *Cities.*, vol. 72, pp. 115-121, 2018. <https://doi.org/10.1016/j.cities.2017.07.010>.
- [51] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *Comput Sci*, vol. 2018, 2018. <https://doi.org/10.48550/arXiv.1810.04805>.