

^{1*} He Fangzhou^{2,3} Bai Wei⁴ Wang Zhiqi

Knowledge Graph Construction and Digital Twin Modeling Integrating Multi-modal Data



Abstract: - Supply chain analytics is focusing more and more attention on industrial logistics. Optimal allocation of assets is severely restricted by the timing inconsistency and geographical instability of industrial logistics assets brought by unpredictability and variability. Unnecessarily lengthy drive distances and lengthy delays are caused by the incapacity to acquire and utilize industrial logistics asset spatial-and-temporal (ST) qualities rationally, which affects the processes' potential to operate sustainably. Thus, for efficient usage of resources in industrial logistics, a novel machine learning (ML)-assisted knowledge graph (ML-KG) design is provided in this work. For evaluating the multi-modal data produced by widely deployed IoT devices, a novel customizable diversified kernelized support vector machine (CDK-SVM) approach is also suggested. After establishing the suggested ML-KG framework for ST integrity in the representation of the digital twin version, links are carried out and reasoning is performed using job data related to industrial logistics. The graph mechanism effectively distributes the industrial logistics resources. Lastly, the outcome proves that the suggested process is successful in the distribution of industrial logistics assets.

Keywords: Industrial logistics, knowledge graph, digital twin, usage of resources, machine learning (ML)-assisted knowledge graph (ML-KG), customizable diversified kernelized support vector machine (CDK-SVM)

I. INTRODUCTION

In the age of the digital era, the crossbreeding of diverse technologies has set off the conceptual chain reaction in the human grasping of the systems' complexity [1]. As these, knowledge graph construction and digital twin modelling become the leading methodologies that make it possible to create a complete representation and the construction of a simulacrum by reference to real-world entities. Work on making these technologies increasingly valuable, since they can not only combine disparate data sources but also provide unprecedented insights into smart cities, manufacturing, healthcare, along many other fields [2].

That is the semantic core of knowledge graph-building – assembling connected networks consisting of the respective entities and their attributes that are characterized as structured information using the semantically rich format [3]. Through this structured representation, searching and retrieving the data as well as analyzing data become quick and easy, so the knowledge becomes deeper about the complex relationships in the field. Data processing and advanced analysis capacities have their peak with knowledge graphs, which assist in one-of-a-kind extraction and transformation of data into useful and applicable knowledge from immense and diverse datasets [4].

The incorporation of varied data modalities into both knowledge graph constructions and a digital twin model brings additional dimensions to the analysis and helps to characterize more completely the complexities of the systems that are examined [5]. Multi-modal data sets the complexity and diversity of different types of information which includes textual, numeric, spatial, temporal and sensor-derived information, each providing a singular level into the different areas of the system and environment [6].

As a current example, in the smart city domain where the data from sources such as traffic sensors, weather forecasts, social media feeds and urban infrastructure databases is integrated, the probability for the creation of dynamic knowledge graphs that will capture the city in real-time becomes high [7]. The slight overlaying of these knowledge graphs with digital twin models of the strategic infrastructure components' systems will enable planners to perform various scenarios simulations, evaluate the intervention effects and enhance urban operations regarding live ability, sustainability and resilience [8].

In the same way, the infrastructure of industrial manufacturing systems can be upgraded to allow streaming of different data from devices, supply chain databases and production logs which can, in turn, provide a basis for the development of the complex models that replicate the entire plants [9]. Tying digital replicas of machines with material flows and quality-related standards, as well as with manufacturing processes knowledge graphs enables

¹ *Corresponding author: He Fangzhou, housel@qq.com, School of Public Security Information Technology and Intelligence, Criminal Investigation Police University of China, Shenyang, 110854, China

² Intelligent Policing Key Laboratory of Sichuan Province, Luzhou, 646000, China

³ Department of Transportation Management, Sichuan Police College, Luzhou, 646000, China

⁴ School of Investigation and Counter-Terrorism, Criminal Investigation Police University of China, Shenyang, 110854, China

Copyright © JES 2024 on-line : journal.esrgroups.org

manufacturers to be predictive in maintenance, provide scheduling optimization and minimize risks of failure or faulty products [10]. Despite its promise, knowledge graph and digital twin models confront difficulties with data integration, scalability and interoperability [11]. It might be difficult to coordinate disparate data sources and maintain consistency across changing systems. Furthermore, guaranteeing the security and privacy of sensitive data inside networked models remains a major challenge.

The development of digital twin technology has created huge public interest, while there are evident logistical manufacturing obstacles. The large amount of data constituting IoT sensors with spatial-temporal value could result in environmental basics if it is not properly exploited. The main problem is the way of converting the dynamic data from IoT systems into a structured map with time information. Currently, the digital twin researches, to a small amount, are focused on single elements or processes, but the eternity and the links of time and space are separated although the phenomenon is taking place on the same matter. Along with this is forming the right type of spatial-temporal model that has the detailed features of the global production logistics resources. The production logistics of allocation to resources are yet under developed as well. Efficient resource allocation is, therefore, all about thinking in advance and, mainly, about taking into account resource dynamics in the framework of activities. Thus, it is necessary to conduct broad-scale studies providing academic material distribution to cope with spatial-temporal dynamics as well as operational monitoring and control. This limitation motivates to solve the difficulty by providing a relational spatial-temporal knowledge graph model that enables optimal IL resources management. The model is initiated by defining the spatial-time variables by a customizable diversified kernelized support vector machine (CDK-SVM) built to use that specific parameter. Lastly, the knowledge graph is a dynamic one that encompasses all valuable information resources and connections that are involved. We design a resource allocation method (graph algorithm) that considers not only the geospatial distance bearing but also the constraints of time. A use case in real life in which we are to show the effectiveness of the proposed strategy could be applied.

II. LITERATURE REVIEW

We performed a thorough analysis of the most recent research on knowledge graphs, digital twin technologies and production logistics resource management in the manufacturing industry, with an emphasis on resolving the relevant issues involving such areas, table 1 illustrates the traditional study of digital twin, knowledge graph and industrial logistic.

A. *Manufacturing knowledge graph*

The goal of the research created a knowledge graph for manufacturing resources in cloud manufacturing settings by using a knowledge extraction and fusion technique [12]. It created an ontology model based on resource features and presented an interactive visualization analysis approach to validate resource-finding apps.

They utilize multi-agent reinforcement learning and an industrial knowledge graph to create a Self-X cognitive industrial network [13]. While acknowledging its limits and urging more study for improved smart manufacturing systems, it shows encouraging results in accomplishing semantic-based self-configuration and decentralized self-optimization.

The framework outlined in article proposed a connective framework and an ontology-based MK graph to solve fragmented knowledge reuse in manufacturing, resulting in improved problem-solving decision-making [14]. It displays good integration and performance in tackling actual production challenges using semantics-based knowledge computing.

B. *Resource management and logistics for the industry*

Integrated planning model for the supply chain of sugar beets was created in [15], which maximizing industrial and agricultural choices to reduce operating expenses. To accomplish it made use of binary integer programming, recognizing the inherent complexity while showcasing its efficacy through a genuine case study.

An enhanced ant colony algorithm and a fuzzy time window scheduling model were two of the real-time data-driven methods employed in the research [16] to optimize production logistics in dynamic industrial situations. Although there can be some scalability and generalizability difficulties, the results demonstrated greater feasibility and distribution cost reduction in a machining workshop.

By reducing uncertainty using improved VRPDP models and digital twin-driven designs, research [17] was to enhance decision-making in production logistics systems. It offered methods for dealing with dynamic disturbances in real-time synchronization, which were assessed via a case study, highlighting their applicability and utility in such instances.

The purpose of the study [18] addressed the dearth of quantitative methods for evaluating the effectiveness of production logistics caused by various production models. A case study comparing workshop efficiency before and after cellular manufacturing implementation demonstrated the usefulness of the evaluation indexes and fuzzy entropy model that were used to quantify operational efficiency. Applicability to different production situations could be one of the limitations.

Mining anomalies from RFID data in production logistics was a challenging issue that the research paper [19] addressed using a novel approach that made use of clustering algorithms and multi-attribute views. According to experiments, the technology was valuable in improving logistics management and monitoring since it can detect over 90% of irregularities with accuracy. The use of exact RFID data collection techniques could be necessary and larger datasets can offer scalability issues.

A novel model and method for real-time scheduling in dynamic workshops that includes predictive elements and an information updating mechanism was presented in the study [20]. When compared to existing approaches, the results show greater performance in terms of customer satisfaction, equipment usage and energy consumption, while maintaining acceptable schedule resilience. The suggested approach's capacity to scale to extremely large-scale applications could prove limited.

C. Digital twin

A 5-dimensional model was used in the research [21] to examine and define supporting technologies and methodologies for digital twin deployment. The process consisted of investigating and evaluating prevalent techniques. The findings shed light on commonly utilized technology and methodologies, while also acknowledging the intricacies and present challenges of digital twin adoption. Limitations include the changing nature of technology and the need for ongoing adaptation.

Improved in-house logistics by enabling agile decision-making with simulation-based tools was discussed in [22]. It evaluated two models through real-world operations, indicating that they were representative and relevant to a variety of logistical settings, but with possible limits in complicated scenarios.

Using Axiomatic Design theory to maximize deployment strategies, the author in research [23] offered an implementation strategy for the physical component of Digital Twin technology in industrial systems. The strategy's operability and efficacy in improving manufacturing processes were demonstrated through a case study involving CNC machines.

An SDT framework was introduced in research [24] that utilizes five-dimensional modeling, real-time monitoring and predictive analytics to overcome the difficulties associated with integrating Digital Twin (DT) technology during the manufacturing phase. Using an engineering case study, they develop a DT-VMPS for shop-floor operations, noting future research topics and demonstrating the system's practicality and effectiveness.

Digital twin-driven method was utilized by author in [25] to improve energy-efficient multi-crane scheduling while taking unpredictability in crane energy usage into account. The research showed possible energy savings by simulating crane operations, scheduling jobs and evaluating energy use using a DT framework. The realistic applicability and accuracy of the simulation might have limitations.

Table 1: Summary of literature review main objective, method, advantage and disadvantage

Reference	Objective	Method	Advantage	Disadvantage
[12]	In cloud manufacturing contexts, to build a knowledge graph for manufacturing resources.	Using an ontology model, created an interactive visualization analysis technique and used a system for knowledge extraction and fusion.	Effective resource identification, easy access to resources for processing tasks, and assistance with the creation of resource pools for cloud manufacturing.	Possibility for ontology model development complexity, reliance on precise knowledge extraction, and scalability issues.
[13]	Utilizing MARL and IKG, create a Self-X cognitive manufacture network.	Create Interactive Knowledge Graph (IKG), use graph neural network-based embedding, and apply Multi-Agent Reinforcement Learning (MARL) into operation.	Realizes decentralized self-optimization and semantic-based self-configuration	Self-X levels had limited preparedness; further research was needed to improve capabilities.

[14]	Used the connectivism paradigm and the ontology-based Manufacturing Knowledge(MK) graph to address the fragmented reuse of knowledge in manufacturing.	Create semantic-based knowledge computation to improve decision-making while addressing problems.	Provides an all-encompassing method for combining manufacturing expertise with the resolution of production issues.	Demands a substantial upfront outlay of funds for the creation and use of the suggested techniques.
[15]	Create an integrated supply chain management model for sugar beets.	Create a sugar beet supply chain management model.	Reduces operating costs, incorporates crop rotation, and takes spatial and temporal differences into effect.	The binary integer programming technique may result in a computational load and complexity in both modeling and solution.
[16]	Utilize real-time, data-driven techniques to optimize p in dynamic industrial environments.	For optimization, use an upgraded ant colony algorithm combined with a fuzzy time window scheduling model.	In real-time production situations, attain enhanced feasibility and cost reduction of distribution.	Scalability issues and restrictions on extrapolating results are examples of potential drawbacks.
[17]	Using digital twins and enhanced VRPDP models, improve production logistics decision-making.	Create real-time synchronization methods, then use a case study to validate them.	Enhances flexibility and decision-making in the face of changing uncertainty	Insufficient resources might make mass adoption more difficult due to implementation difficulty.
[18]	Address the absence of quantitative tools for evaluating manufacturing logistics efficiency.	Establish a fuzzy entropy model and assessment indices to gauge operational efficiency.	Provides a systematic method for assessing production logistics efficiency using quantitative measures.	Could require to be modified because of its limited applicability to various manufacturing scenarios.
[19]	In manufacturing logistics, tackle the problem of identifying irregularities using RFID data.	Offer an innovative approach that takes into consideration clustering methods and multi-attribute perspectives.	Over 90% of abnormalities are efficiently identified, improving logistics management and monitoring.	Dependency on precise RFID data gathering procedures and possible scalability problems with bigger datasets
[20]	Improve the scheduling process in dynamic workshops in real-time.	Introduce a novel algorithm and scheduling model that includes update mechanisms and predictive components.	superior results in terms of energy usage, equipment utilization and customer satisfaction	Possible restrictions on the scalability to extraordinarily large-scale situations
[21]	Investigate enabling technologies for the digital twin.	Analyzed with a 5-dimensional model.	Provides insights on frequently utilized technology for digital twins.	Complexity and continuing technological innovation are two significant challenges.
[22]	Improve internal logistics through rapid decision-making.	Develop and evaluate models using simulation-based techniques.	Offers representative digital-twinning technologies for operational development.	Limitations may exist in complicated logistical settings.
[23]	Develop a strategy for the execution of the physical component of DT in industrial systems.	For deployment strategy optimization, apply the philosophy of Axiomatic Design.	improves production procedures by utilizing DT technology	Limited study on the physical entity element, with possible implementation issues.

[24]	Examine the difficulties in implementing SDT for production using the suggested framework.	Make use of real-time tracking, predictive analytics and five-dimensional modelling.	Increase productivity and predictive skills on the shop floor by utilizing DT-VMPS.	SDT system setup and maintenance might provide some challenges.
[25]	Enhance energy-efficient multi-crane scheduling by taking into account the unpredictability of crane energy use using a digital twin-driven strategy.	Simulate crane operations, plan jobs, assess energy usage, and show possible energy savings by using the DT framework.	Provides dynamic performance evaluation and real-time mapping to improve energy economy and the precision of crane scheduling.	For real-world applications, simulation accuracy was required; nevertheless, the complexity of actual crane operations might not be adequately represented.

III. SPATIAL AND TEMPORAL MAPPING OF DIGITAL TWINS

A. Machine learning-assisted knowledge graph (ML-KG)

Figure 1 depicts the four-layer design of ML-KG-driven industrial logistics (IL) resource allocation, which solves the issues raised above. We've built a cutting-edge IoT infrastructure across our industrial park, including indoor and outdoor areas, to gather actual ST data's using ML. This physical configuration serves as the foundation for our digital operations analysis. Moving further, we intend to create digital twins to map spatial-temporal dynamics, create dynamic knowledge graphs and use graph algorithms to optimize resource allocation for accurate decision-making. These processes study together to thoughtfully address the research problems.

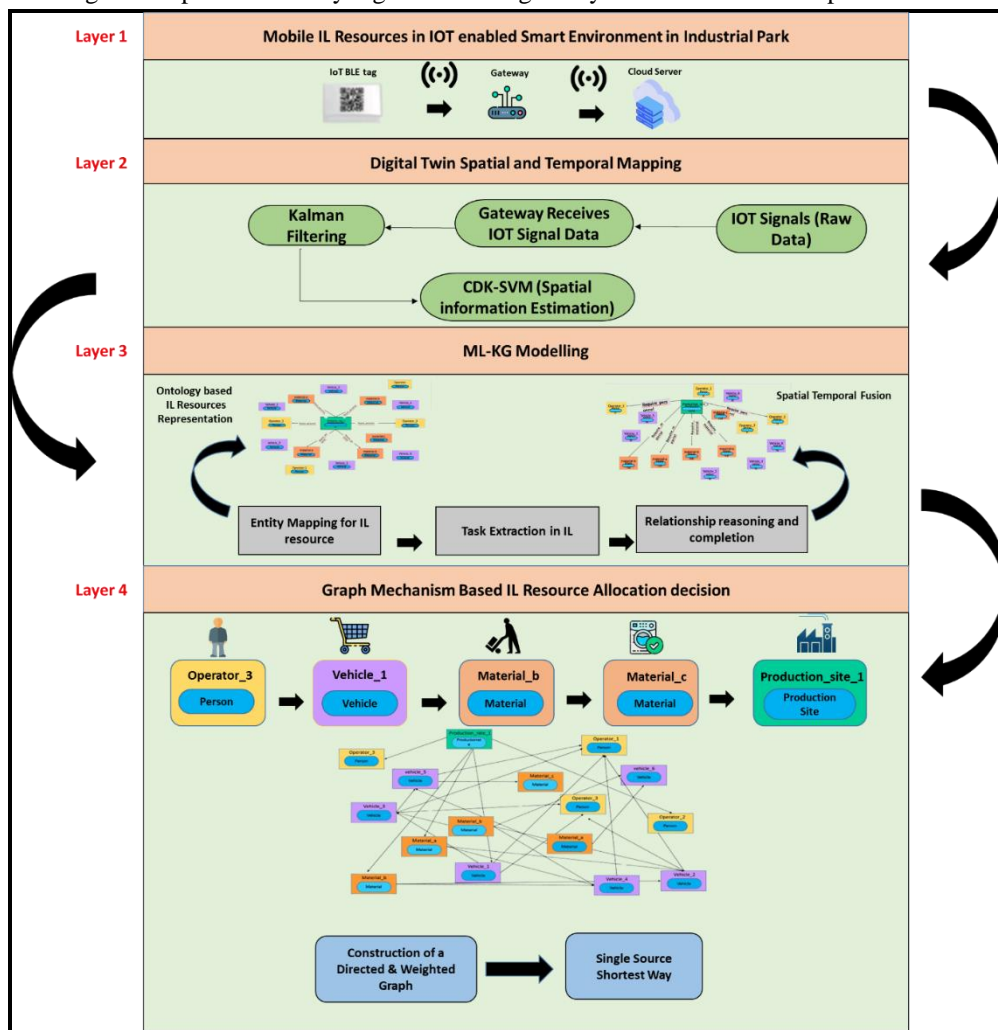


Figure 1: Dynamic Allocation of Resources in Spatial and Temporal Knowledge Graph-based Machine Learning Systems

B. *IOT-Empowered mobile IL resources in smart environments*

To efficiently manage operations in the industrial park, IL activities must be carried out with the continuous flow of resources such as persons, trucks and materials between production lines and workshops. Tracking these resources, however, is difficult due to their frequent movement and unpredictable nature. While the GPS and BDS are effective for outside locations, the majority of IL activities take place indoors over numerous levels, where satellite-based systems fail. To solve this, BLE tags are placed on mobile resources and continually broadcast signals. Gateways strategically placed across the park gather these signals and use edge computing to analyze and filter noisy data, compressing it before passing it to the cloud server. This integrated method provides real-time monitoring and administration of IL resources in indoor as well as outdoor environments.

Objective: This layer is intended to extract geographical and temporal insights from signal data produced by IoT devices. Its goal is to correctly convert the spatial-temporal properties of physical-world resources with the digital domain, hence improving our knowledge of data logistics resources.

CDK-SVM generates cyberspace location data to assure spatial and temporal consistency across physical and virtual worlds. Virtual copies of IL resources having spatial-and-temporal properties are created by the machine learning method, drawing influence from the real-world items in issue. The broadband signal strength of the gateway was noisy and fluctuated due to the impacts of multipath and attenuation. Among the most effective algorithms for de-noising data is the Kalman Filter. The Kalman Filter is used to filter the raw RSSI information from many gateways. To train a non-linear model using supervised learning, datasets are necessary. Signal strength is affected by the physical environment around it, which includes things such as walls, obstacles, humidity and temperature. An associated RSSI value and location labels reflecting local environmental data are among the training datasets that are developed for the signal fingerprint database. A CDK-SVM is built to be efficient as it responds to environmental changes, ensuring the durability and dependability of estimations of geographical information. To calibrate the CDK-SVM's parameters and weights, updated estimations of online spatial information will be resent into the model.

C. *Machine learning-assisted knowledge graph (ML-KG) modeling*

A KG is a different network where the edges represent links between items and the vertices represent subjects. Mapping IL resources and their characteristics, such as spatial and temporal data, enable the creation of a knowledge graph that accurately depicts the geographic distribution and the interactions among IL resources. In the graph, the entities of IL resources are mapped with attributes and the ontology-based knowledge of IL resources is conveyed. The entities build a relationship when pertinent IL resources are taken out of IL tasks. The IL task's logic construction, together with the progressive integration of IL resources, encourages association reasoning and completion inside the KG. The directional knowledge graph is then constructed. The KG graphically represents the resources' spatial-temporal values. Combine geographical and temporal data from the layer below to create a coherent ML-KG. Edges in a KG represent the logical relationship between objects, whereas vertices represent actual locations in the real world. The apparent spatial and temporal consistency between the virtual imitation and the actual thing is achieved through the use of KG modeling.

D. *Decision Making in IL Resource Allocation through Graph Mechanism*

This layer makes decisions based on graph algorithms and allocates IL resources to tasks depending on IL. When determining the weights on the edges produced by the layer below, two limitations were taken into account, time and transit capacity. The knowledge graph's weights are determined using the Euclidean distance, resulting in a directed graph having weighted edges. The cornerstone of IL task extraction lies in recognizing the perspective from the transportation endpoint to the task performer, known as the destination viewpoint. This transforms the challenge of allocating resources into a single-source shortest route problem concerning time constraints and transport capacity constraints. Refocus the task on completing the nearest available resource-utilizing action while prioritizing minimizing travel distance. First, the most efficient path is determined using Dijkstra's shortest route method. This process is then reversed to determine the direction of resource distribution.

IV. MAKING RESOURCE ALLOCATION DECISIONS USING ML-KG ASSISTED BY DIGITAL TWIN

A. *CDK-SVM: To map digital twins spatial-and-temporal*

Because indoor localization resources are located in multi-level interior environments that are inoperable for satellite positioning systems, estimating spatial information about them is a difficult task. The spatial information of real-time IL resources can be estimated using the CDK-SVM that we present in this research. Remarkable

outcomes were demonstrated by indoor positioning systems including, TDOA, TOA and AOA, nevertheless, extra gear is required, which could result in significant costs when used on a big scale. RSSI, which is widely used in wireless devices, serves as a simple measure for determining the intensity of a signal from a transmitter to a receiving device, eliminating the need for extra supporting equipment. In this study, we use RSSI values as the primary characteristics for spatial information estimates.

The performance of Support Vector Machines (SVM) is largely dependent on the kernel function selection, which is impacted by the data distribution characteristics of the feature space. Kernel functions are roughly divided into two types, local and global. Local kernel functions thrive in learning but can struggle with generalization, whereas global kernel functions excel at generalization but have lower learning capabilities. Common examples include the Gaussian radial basis kernel function (RBF) for the local type and the polynomial kernel function (Poly) for the global type. These kernels are often used in traffic flow prediction applications. Their respective equations are as follows:

$$m(w, w_i) = [\gamma(w * w_i) + 1]^q \tag{1}$$

$$m(w, w) = \exp(-\gamma\|w - w_i\|^2) \tag{2}$$

Poly is represented by Eq. (1), and RBF by Eq. (2). The kernel function, γ , determines how data is distributed after mapping to a new feature space, whereas q denotes power.

$$m(w, w_i) = \beta \cdot \exp(-\gamma\|w - w_i\|^2) + (1 - \beta) \cdot [\gamma(w * w_i) + 1]^q \tag{3}$$

$\beta \in [0,1]$ represents the weight coefficient of the hybrid kernel function in Eq. (3).

When $m = \frac{u_{i-1} - u_{i-2}}{u_{i-1} - u_{i-2}}$, It reflects the slope of the latest two data points for traffic volume. As the absolute value of k declines, the traffic flow curve becomes smoother. To improve the model's global generalizability, we can either raise or reduce the relevance of the polynomial kernel function (β). Conversely, as the absolute value of k grows, the curve tends to sharpen. Increasing the value of β or the weight of the function known as the Gaussian kernel can improve the capacity of the model for local learning.

B. Decision-making for resource distribution using ML-KG models

The procedures for allocating IL resources using the ML-KG are displayed in Figure 2. The knowledge graph initially represents the entity of IL resources as vertices with different categories, encompassing their basic properties. Personnel expertise spans a wide range of positions, including operators experienced in specific machinery operations, driver's adept in vehicle operation and technicians with specialized talents. These workers are responsible for moving products utilizing vehicles, which might include WIP, components, materials and tools that belong within the product category.

The site group, which includes buffers and stations, serves as the start and finish point for IL procedures. The property layer's characteristics are associated with each item in the knowledge network to explain restrictions, execute spatial-temporal calculations and make determinations. Every IL mission includes thematic information regarding transporting personnel/products to a certain place. The site entity's task requirement attribute contains unstructured IL task data. The module-based connection extraction approach is used to generate directed linkages from the source to the target KG entities. The knowledge graph's core edges and vertices are well-defined. Section 3's spatial-temporal estimate values are subsequently mapped to graph entities. The entity's location and timestamps are updated in real time on the graph databases and visualized on the spatial system, as illustrated in Figure 2.

Finding the most economical resource to carry out a sequence of IL activities limited by capacity and time window using geographical data is the aim of the IL resource allocation process. Therefore, given certain limitations, we convert the IL logistics resource allocation issue into the single-source shortest pathfinding problem. Since all of the weights in the graph algorithm are nonnegative, Dijkstra is chosen. For each vertex in our set R, the distance is pre-established and accurate. Algorithm 1 determines the shortest path between the site entity, indicated as S and the final command unit. First, create the appropriate data structures by making the previous vertex on the optimal path from sources undefined and assigning a starting distance of infinite to all succeeding vertices. Q denotes all un-optimized vertices acquired by ML-KG. During each iteration, the method relaxes the exterior edges of all vertices, chooses the vertex outside R with the shortest distance from it, as well as inserts it into R. This procedure will continue until the shortest path is found. The decision about the location of the nearest vertex to the visitor is critical. Once recognized, each of its nearby vertices is removed from Q, and relaxation is applied accordingly. To determine if the moment of arrival falls within a specific time range, divide the Euclidean distance by the IL resource's velocity to calculate the duration. To effectively record and compute the best path, a variety of data structures such as Prev [], Dist [] and others are used. The knowledge graph's primary edges and vertices are preset, allowing for more efficient operations. The execution viewpoint with the appropriate cost is the

opposite of the optimum approach. Building the initial priority queue for $|V|$ takes time. All of the ML-KG vertices can be explored through edges in time with the adjacency list. Since the vertex is deleted from Q once every loop, the loop's iterates. The minimal vertex is updated after that is taken out of Q .

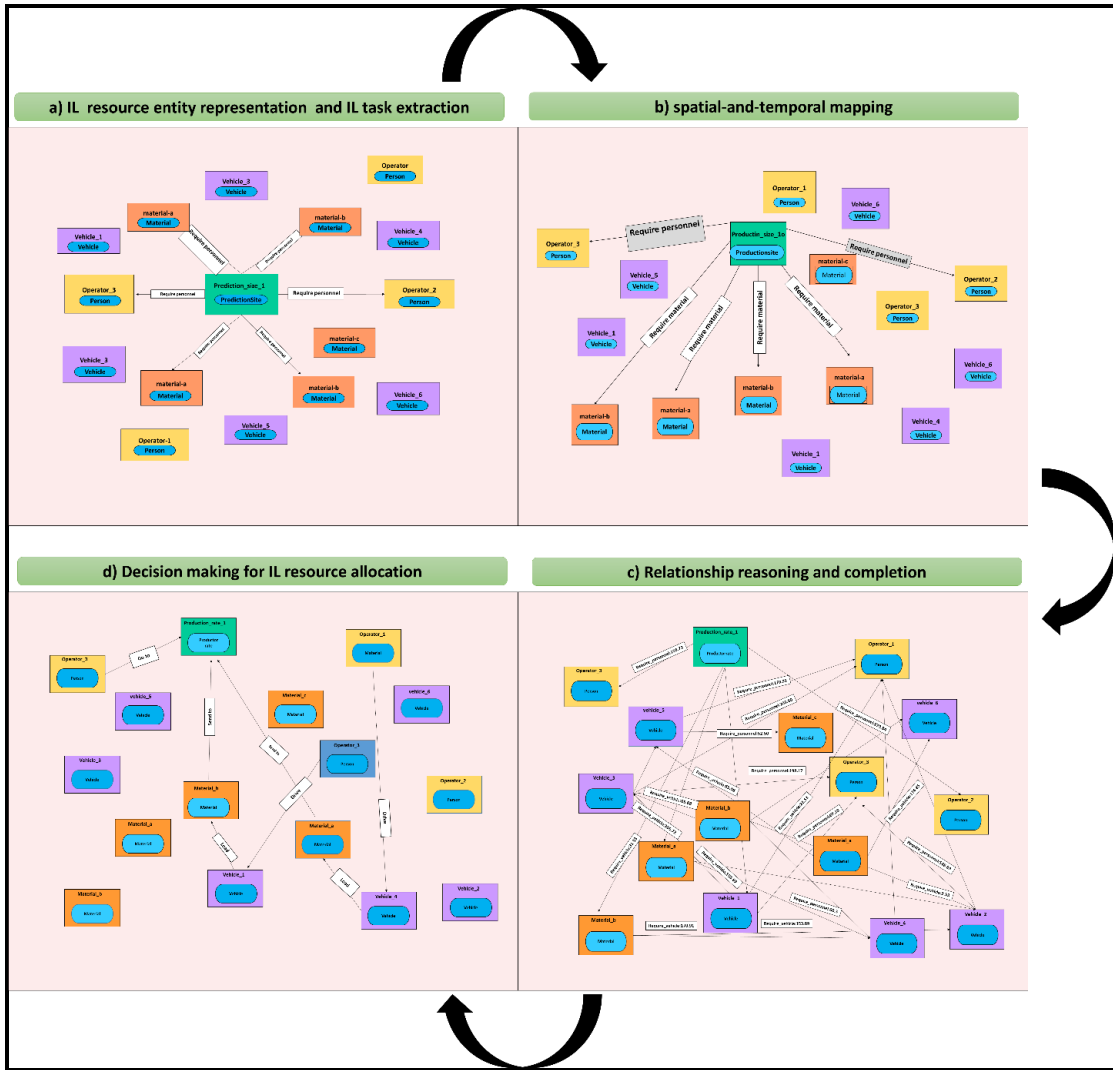


Figure 2: Process of dynamically allocating resources using spatial and temporal KG

Algorithm 1: Process of ML-KG:

Input: $ML-KG (V, E), S$

Output: $prev[], dist[]$

- I. $Dis[u] \leftarrow \infty, prev[u] \leftarrow nil$
- II. $Dist[S]$
- III. $Q \leftarrow$ the collection of all vertices that are optimized
- IV. while $Q \neq \emptyset$:
- V. $u \leftarrow$ The lowest distance in Q is the vertex u
- VI. For all $(u, v) \in E$:
- VII. If $dis[v] > dist[u] + w(u, v)$ AND
- VIII. $arrTime[v] \in S.timeWindow$ AND
- IX. $Capacity[v] \leq S.capacity\ requirement$
- X. $dist[v] \leftarrow dist[u] + w(u, v)$
- XI. $prev[v] \leftarrow u$
- XII. update Q
- XIII. Return $prev[], dist[]$

C. Scenario

A company specializing in manufacturing washing machines is facing challenges in efficiently monitoring and allocating resources, leading to decreased productivity. Their operations span residential washing machines, intelligent equipment and home appliances. To enhance collaboration and reduce delivery costs, they've established an industrial park where various suppliers, including upstream and downstream partners, are invited to operate. The park houses multilevel plants for central washing machines, evaporators and condensers, equipped with assembly lines and machinery. Outdoor and semi-outdoor areas are designated for parking and injection, ensuring proper ventilation. However, the complex spatial-temporal dynamics of indoor, outdoor and semi-indoor environments pose challenges for resource management. Personnel, material carts, forklifts and merchandise are spread around the park. To overcome these difficulties, the organization needs an immediate digitization and intelligent transformation solution. To assess the success of their suggested solution, the study team developed a full software and hardware system in the collaboration firm.

D. ML-KG deployments

Real-time spatial-temporal data collection and filtering must first be done in an IoT smart environment before ML-KG can be used. For big-scale tracking, the research team investigates several technologies. "Large-scale" means tracking many things and covering a vast area of floor, which might provide maintenance issues if smart tech battery life is limited. Timing, expense and precision are crucial in this case. Offering minimal tag/gateway costs, up to three years of operation and meter-level accuracy, BLE stands out as a compromise.

As indicated in the placement cell, the gateway is fixedly positioned close to the power supply. Forklifts, employees and material trolleys are just a few of the mobile IL resources that are connected to different BLE tags. BLE tags are carried by 234 employees, 522 vehicles and 38 different kinds of materials. There are 98 gates located both indoors and outside. Signal collecting using the typical fingerprinting signal collection strategy (TFSCS) requires a lot of time and effort since it splits monitoring regions into uniform location cells. Scalable location cell signal collection (SLCSC) is used as a countermeasure. To reduce effort, it dynamically modifies cell size. Cell size is reduced for accurate placement in high-accuracy locations, such as interior multi-layer situations. However, cell size increases in low-accuracy environments, such as outdoor regions, to minimize signal-gathering efforts.

E. Performance analysis

The industrial park uses 98 gateways. Input neurons correspond to gateway counts, resulting in input vector $\hat{x} = [r_1, r_2, r_3, \dots, r_{98}]^T$ where the RSSI value obtained via the i -th gateway is represented by r^i . Unreachable gates ways in a -110 value for constructing the input vector.

Signal collection time in an illustration site refers to the timeframe when data is gathered from numerous sources or sensors at a single site for analysis or monitoring reasons. Figure 3 depicts the performance statistics for ML-KG. Initially, we track the IL resources before assigning them using the graph method. In the data set preparation phase, signal collection time is especially decreased in SLCSC compared to traditional fingerprinting signal collection strategy (TFOSC), particularly at FBA (s2), IBW (s3) and ECP (s4), owing to reduced accuracy demands.

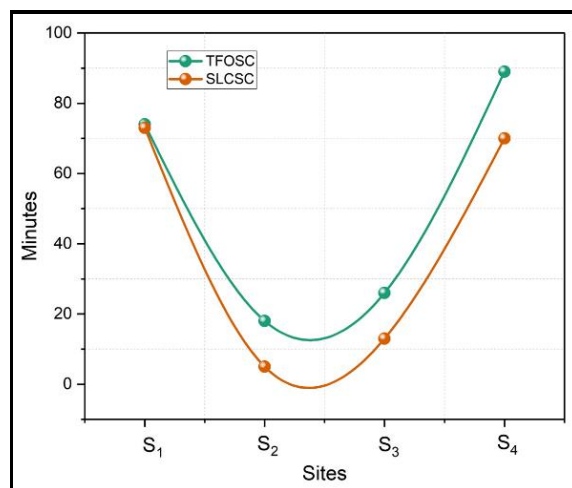


Figure 3: Signal collection time

Vehicle utilization rate is the ratio of the time a vehicle is actively employed for transportation to its total available time. Figure 4 shows a significant increase in vehicle utilization due to the ML-KG's capacity to locate each vehicle's location and integrate the data into the process of making decisions.

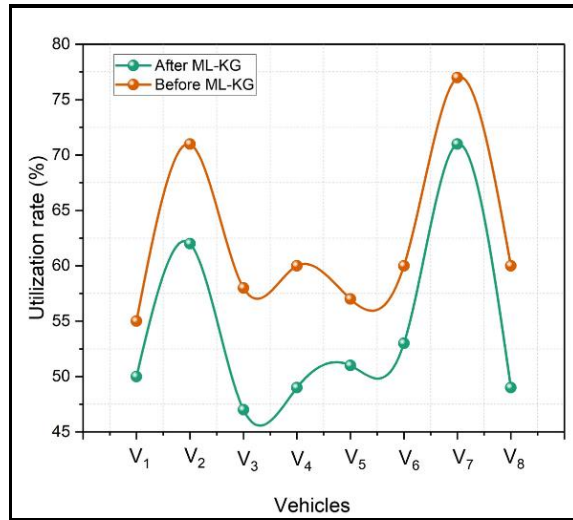


Figure 4: Vehicle utilization rate comparison

Waiting time is the amount of time that a unit or item is inactive or in a queue before that is processed or transferred to the following stage of a workflow or manufacturing process. It refers to the time when resources are not actively engaged in value-adding tasks, which can result in inefficiencies and delays. Figure 5 clearly shows the considerable reduction in wait times at both the logistics and M/WIP buffers before as well as after ML-KG implementation. This reduction in waiting times demonstrates a considerable increase in temporal synchrony throughout the site's logistical activities.

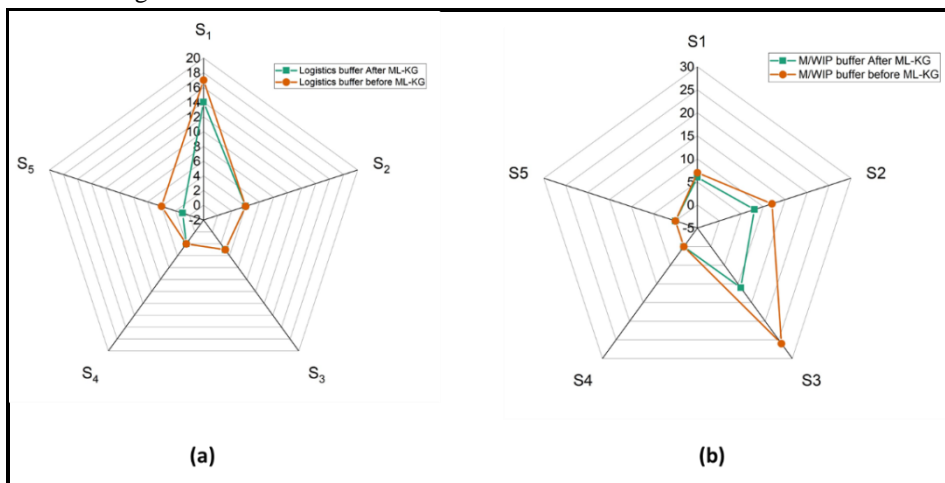


Figure 5: Typical IL waiting time

The average traveling distance task order in industrial logistics refers to the normal distance products travel from origin to destination in a logistical process, which is frequently quantified in kilometres. Significant reductions in average distance travelled by various resources for the IL task were accomplished by developing a Dijkstra-based graph algorithm, thereby improving the way planning process, Figure 6.

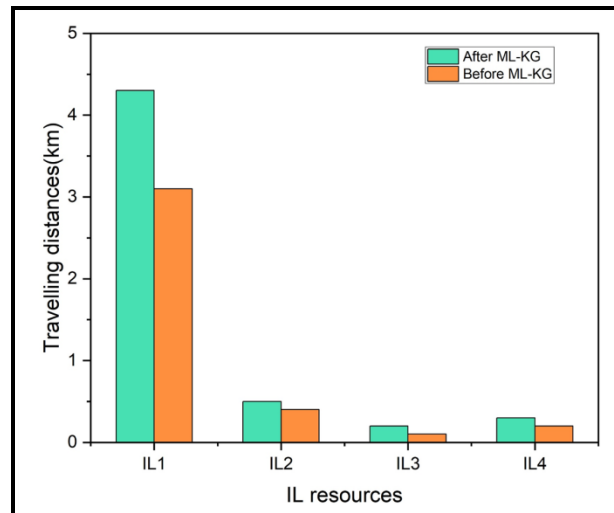


Figure 6: Average traveling distance per IL order task (km)

V. CONCLUSION

This study focuses on the issues of efficient resource allocation in complex industrial contexts where spatial disorder and temporal asynchrony of resources impede sustainable development. Focusing on a real-world industrial park, we offer a strategy that begins with mapping spatial-temporal resource values using CDK-SVM and IoT sensor data. Then, our ML-KG model extracts information and relationships from IL procedures, generating virtual entities suitable for visualization. To represent these items, we construct a directed and weighted graph that incorporates edge reasoning and geolocation information. Finally, we offer a Dijkstra-based graph method for optimizing resource allocation, taking into account restrictions such as time windows and capacity, to reduce overall travel distance. This study enhances both the scientific understanding of digital twins in IL and the practical knowledge of resource allocation in the industrial sector.

ACKNOWLEDGMENT

This research was funded by Intelligent Policing Key Laboratory of Sichuan Province Project (No.ZNJW2022KFMS002); Liaoning Province social science planning fund project (No.L21ASH004); Liaoning economic and social development project (No. 2022lslybkt-039); 2023 Liaoning Province Applied Basic Research Program Project (No. 2023JH2/101300150); Shanghai Institute of Criminal Science and Technology Key laboratory of on-site material evidence open project (No. 2019XCWZK06, 2020XCWZK02, 2020XCWZK03).

REFERENCES

- [1] Cao, L., 2023. Trans-AI/DS: transformative, transdisciplinary, and translational artificial intelligence and data science. *International Journal of Data Science and Analytics*, pp.1-13.
- [2] Bibri, S.E., 2021. Data-driven smart sustainable cities of the future: An evidence synthesis approach to a comprehensive state-of-the-art literature review. *Sustainable Futures*, 3, p.100047.
- [3] Zeng, J., Chua, Z.L., Chen, Y., Ji, K., Liang, Z. and Mao, J., 2021, February. WATSON: Abstracting Behaviors from Audit Logs via Aggregation of Contextual Semantics. In *NDSS*.
- [4] Kalaycı, T.E., Bricelj, B., Lah, M., Pichler, F., Scharrer, M.K. and Rubeša-Zrim, J., 2021. A knowledge graph-based data integration framework applied to battery data management. *Sustainability*, 13(3), p.1583.
- [5] Huang, X., Liu, Y., Huang, L., Onstein, E. and Merschbrock, C., 2023. BIM and IoT data fusion: The data process model perspective. *Automation in Construction*, 149, p.104792.
- [6] Lei, Y., Wang, Z., Chen, F., Wang, G., Wang, P. and Yang, Y., 2023. Recent Advances in Multi-modal 3D Scene Understanding: A Comprehensive Survey and Evaluation. *arXiv preprint arXiv:2310.15676*.
- [7] Qi, L., Li, J., Wang, Y. and Gao, X., 2019. Urban observation: Integration of remote sensing and social media data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(11), pp.4252-4264.
- [8] Gallotti, R., Sacco, P. and De Domenico, M., 2021. Complex urban systems: challenges and integrated solutions for the sustainability and resilience of cities. *Complexity*, 2021, pp.1-15.
- [9] Bicocchi, N., Cabri, G., Mandreoli, F. and Mecella, M., 2019. Dynamic digital factories for agile supply chains: An architectural approach. *Journal of Industrial Information Integration*, 15, pp.111-121.

- [10] Zonta, T., da Costa, C.A., Zeiser, F.A., de Oliveira Ramos, G., Kunst, R. and da Rosa Righi, R., 2022. A predictive maintenance model for optimizing production schedules using deep neural networks. *Journal of Manufacturing Systems*, 62, pp.450-462.
- [11] Hemdan, E.E.D., El-Shafai, W. and Sayed, A., 2023. Integrating digital twins with IoT-based blockchain: concept, architecture, challenges, and future scope. *Wireless Personal Communications*, 131(3), pp.2193-2216.
- [12] Yin, C., Li, B. and Chen, J.X., 2021, June. The construction method of knowledge graph of manufacturing resources in a cloud manufacturing environment. In *2021 IEEE 4th International Conference on Nanoscience and Technology (ICNST)* (pp. 28-35). IEEE.
- [13] Zheng, P., Xia, L., Li, C., Li, X., and Liu, B., 2021. Towards Self-X cognitive manufacturing network: An industrial knowledge graph-based multi-agent reinforcement learning approach. *Journal of Manufacturing Systems*, 61, pp.16-26.
- [14] He, L. and Jiang, P., 2019. Manufacturing knowledge graph: a connectivism to answer production problems query with knowledge reuse. *IEEE Access*, 7, pp.101231-101244.
- [15] Fikry, I., Gheith, M. and Eltawil, A., 2021. An integrated production-logistics-crop rotation planning model for sugar beet supply chains. *Computers & Industrial Engineering*, 157, p.107300.
- [16] Huang, S., Guo, Y., Zha, S. and Wang, Y., 2019. An internet-of-things-based production logistics optimization method for discrete manufacturing. *International Journal of Computer Integrated Manufacturing*, 32(1), pp.13-26.
- [17] Pan, Y.H., Wu, N.Q., Qu, T., Li, P.Z., Zhang, K. and Guo, H.F., 2021. Digital-twin-driven production logistics synchronization system for vehicle routing problems with pick-up and delivery in industrial parks. *International Journal of Computer Integrated Manufacturing*, 34(7-8), pp.814-828.
- [18] Zhang, Z. and David, J., 2019. An entropy-based approach for assessing the operation of production logistics. *Expert Systems with Applications*, 119, pp.118-127.
- [19] Cao, X., Li, T. and Wang, Q., 2019. RFID-based multi-attribute logistics information processing and anomaly mining in production logistics. *International Journal of Production Research*, 57(17), pp.5453-5466.
- [20] Cai, L., Li, W., Luo, Y. and He, L., 2023. Real-time scheduling simulation optimization of job shop in a production-logistics collaborative environment. *International Journal of Production Research*, 61(5), pp.1373-1393.
- [21] Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., Wang, L. and Nee, A.Y., 2021. Enabling technologies and tools for digital twins. *Journal of Manufacturing Systems*, 58, pp.3-21.
- [22] Coelho, F., Relvas, S. and Barbosa-Póvoa, A.P., 2021. A simulation-based decision support tool for in-house logistics: the basis for a digital twin. *Computers & Industrial Engineering*, 153, p.107094.
- [23] Wei, Y., Hu, T., Wang, Y., Wei, S. and Luo, W., 2022. Implementation strategy of physical entity for manufacturing system digital twin. *Robotics and Computer-Integrated Manufacturing*, 73, p.102259.
- [24] Zhuang, C., Miao, T., Liu, J., and Xiong, H., 2021. The connotation of digital twin, and the construction and application method of shop-floor digital twin. *Robotics and Computer-Integrated Manufacturing*, 68, p.102075.
- [25] Zhao, N., Fu, Z., Sun, Y., Pu, X., and Luo, L., 2022. Digital-twin-driven energy-efficient multi-crane scheduling and crane number selection in workshops. *Journal of Cleaner Production*, 336, p.130175.

APPENDIX -I

RSSI	received signal strength indicator
ST	Spatial-and-temporal
AOA	Angle of Arrival
TDOA	Time Difference of Arrival
TOA	Time of Arrival
WIP	Work-in-process
FBA	Freeboard Adjustment
IBW:	Ideal Body Weight
ECP	Estimated Contract Price
LW	Load Weight
GPS	global positioning system
BDS	BeiDou Navigation Satellite System
BLE	Bluetooth Low Energy
DT-VMPS	Digital twin-based visual monitoring and prediction system
RFID	Radio Frequency Identification
VRPDP	Virtual Reality Product Design and Presentation
SDT	System Dynamics Theory