ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20241209002176
Original scientific paper

Research on Smart City Governance and Network Management Based on Multi-Source Heterogeneous Data Fusion Processing and Analysis

Hang LI, Xiao XIAO*

Abstract: The digital transformation of smart city governance is an inevitable requirement for improving the urban governance system and enhancing the modernization of governance capacity in the context of the development of modern information technology. Firstly, the semantic heterogeneity and syntactic heterogeneity of multi-source heterogeneous data are solved through the unified ontology description of multi-sensor data, and the update algorithm of sensor ontology instance, attribute fusion and feature extraction algorithm are proposed. Secondly, the ontology mapping method is adopted. Aiming at the uncertainty characteristic of multi-source heterogeneous data fusion in smart city, a decision level fusion method is proposed. A transferable data model is used to fuse multiple identification frame information to generate smart city single network management situation assessment results. Finally, the operation of urban "one network management" requires the empowerment of data and technology, from the empirical judgment type to the data analysis type, from the passive disposition type to the active discovery type, to strengthen the coordination and cooperation between departments, from decentralized governance to the overall governance.

Keywords: fusion processing; multi-source heterogeneous data; one network management; smart city; transitive data model

1 INTRODUCTION

The basic connotation of "one network management" is to incorporate as many urban management elements as possible such as things, events, people, organizations (legal persons) into a cross-regional, cross-departmental, online and offline integration of urban management network, and then carry out unified and efficient management of these elements [1]. "One network management" was born in the context of implementing the refined strategy of urban management [2]. District, street and fourth-level cities operate the "one network management" application system to promote the "one network management" in urban management, emergency command, comprehensive law enforcement and other fields. Realize the whole process supervision of city operation situation awareness, physical signs and indicators monitoring, unified event acceptance, intelligent scheduling and command, linkage and collaborative disposal, supervision, evaluation and assessment.

The basic principle of multi-source heterogeneous data fusion technology is to form a consistent description of monitoring objects or a consistent interpretation of monitoring events through rational fusion reasoning of multi-sensor data and control and use in combination with specific application practices [3]. The goal of multi-sensor data fusion is to improve the effectiveness of sensor data fusion system by optimizing the combination of multi-sensor data, so as to discover more effective information about the sensor monitoring target. Multi-source heterogeneous big data fusion technology has been developing rapidly since its inception [4], mainly because it uses multiple different types of sensors for data acquisition, which not only obtains more reliable data than a single sensor, but also adds complementarity to the data obtained from multi-source sensors.

As an important innovation model for the new smart city to promote the modernization of urban governance system and governance capacity, "one network management" has gradually landed projects in many provinces and counties and played an important role since it was proposed. Shanghai, Guangdong, Beijing and other places have launched practical actions of "one network management", opening up the data of multiple departments in the intelligent brain stage in the vertical field. However, these data and intelligent algorithms are still used to serve the business of a single department, such as traffic brain, environmental brain, and so on. Urban governance network management, that is, the modernization of smart city network management after multi-source heterogeneous data integration, is a new generation of infrastructure for urban management, structured on the existing vertical business system of various departments, and fully opens up the smart city network management network (and system) after the integration of multi-source heterogeneous data such as urban management, emergency response, and comprehensive management. Upward can provide auxiliary decision-making for the municipal party committee and municipal government leaders; down can connect communities, streets, support grassroots governance. The two-network integration stage of one network office and one network management, that is, for the whole city area, to "government office one network cooperation" to drive the "dual network" integration of "urban governance one network management" and "government services one network office", to achieve the government and residents for urban governance, co-governance, sharing, and seamless connection of governance and services.

2 RELATED WORK

According to the way of data fusion, multi-source heterogeneous information fusion can be divided into three levels: data fusion, feature fusion and decision fusion [5]. The first is data layer fusion. It is the lowest level of fusion, which directly uses the data monitored by the sensor for preliminary analysis and integration, and then extracts the features of the fusion data to illustrate the time-history characteristics of the engineering structure. Feature layer fusion is relatively simple and flexible, and is more practical in many fusion applications [6]. Third, before the correlation processing of the feature vectors of the observation indicators, it is necessary to describe the observation targets of various types of sensors through pattern recognition [6, 7]. In concrete engineering practice, the first and second fusion modes are often used. By using the back propagation (BP) neural [8] network fusion technology, the displacement monitoring data of several different monitoring points were fused to obtain the comprehensive displacement of the slope. Based on the displacement data after fusion, the mechanical parameters of slope soil are obtained by feedback analysis method. The multi-sensor valuation fusion theory was applied to the dynamic deformation monitoring and analysis of a landslide in southwest China [9], which proved the effectiveness and feasibility of the method in the dynamic deformation monitoring and analysis of landslide. A neural network-based multi-source heterogeneous monitoring data fusion algorithm [10] was proposed, with humidity, wind, cloud cover, daily precipitation and accumulated precipitation as input variables, landslide displacement change data as output data, and measured data were used to verify that [11] network data fusion algorithm was suitable for landslide deformation prediction with multi-source heterogeneous monitoring data. The basic data of dam diagnosis and defect information of on-site inspection were integrated and analyzed [12], and an intelligent dam safety diagnosis system based on monitoring, on-site inspection and hidden historical defects was developed to improve the accuracy of dam safety diagnosis. A multi-sensor observation fusion scheme based on 3D variational data assimilation was developed [13] to integrate GPS and borehole inclinometer data to predict landslide. The results showed that the assimilation process could monitor slope failure more accurately than single GPS data or borehole inclinometer data. A dual-layer heterogeneous sensor network [14] (i.e. geological sensor and camera sensor) was developed to monitor landslides. Once the geological sensor in operation detects a slope anomaly at the first monitoring layer, the camera sensor visually analyzes it, including interlayer triggering, motion detection, and image compression transmission.

The theory of possibility. Based on fuzzy set theory, a possibility theory for event uncertainty is created [15]. Directly inspired by fuzzy set theory, which deals with inaccuracy, possibility theory may compete with probability theory in dealing with uncertainty. The concepts in these two theories are very similar, but the details of the definition differ, such as the confidence function theory [16]. The theory of the confidence function can include many specific theoretical methods, but in fact these specific theoretical methods are very similar to each other, except that under certain conditions the various confidence methods try to obtain certain details. One of the obvious shortcomings of credibility theory is that it lacks a rigorously defined method for interpreting real-world values. The development of credibility theory is still improving. Although it is not widely used in the industrial field [17], it is a very well understood theory in the academic field and can meet certain needs in the reasoning of incomplete information. The data source layer, computing layer, data layer and analysis layer of the smart city after the integration of aviation multi-source heterogeneous data are designed and implemented to provide support for the scientific decision-making of airlines [18]. The fuzzy neural network is applied to the smart city fusion technology [19] after the fusion of multi-source heterogeneous data to obtain the effect of real-time monitoring of gas changes [19]. The research on algorithms has also changed from simple data integration to the application of deep learning [20, 21]. A new smart city fusion method based on multi-source heterogeneous data fusion based on set pair analysis of connection degree is proposed by forming connection degree matrix and expanding dimension [22]. Although China attaches great importance to the development of this aspect, there is still a long distance between it and the international level. Therefore, in order to narrow the gap in computing accuracy and speed of data fusion, further efforts are needed. We will work hard to develop and progress science and technology.

In the research on technological innovation of smart city-network managed cities, the key technology of intelligent holographic mapping for serving smart city-network managed city construction is proposed [23]. Put forward the technical framework of natural scene construction for smart city-network management city construction [24]: Put forward the virtual and real registration method of smart city-network management scene with structural semantic auxiliary constraints, and realize the fusion registration of visual sensor images in virtual scene space [25]. The development of technology has provided strong support for the construction of smart city, but how to effectively integrate and apply them to urban management to achieve true smart city goals remains a challenge. Research on the urban construction mechanism of smart city network management; it is believed that the future-oriented smart city-network management city construction should be promoted from three aspects: digital infrastructure construction, city-level data asset management system construction and new business scenario implementation [26]. It is believed that smart city-network management city construction is faced with the challenges of data, basic knowledge base, multi-system integration and talent problems. The institutional and technical paths to solve these problems are proposed from the macro and micro levels [27]. The city management model of social impact bonds [28] is used to build a collaborative innovation mechanism jointly built by multiple parties. From the four dimensions of knowledge bearing system, knowledge theory system, knowledge algorithm system and knowledge sharing system [29], the method system of urban smart governance driven by smart city one-network management is reconstructed. It is believed that the smart city network management city can understand the operation law of the city, build a closed-loop governance chain, promote the flexible use of resources, and practice the people-oriented concept, so as to comprehensively empower the city "full cycle management" [30]. It is proposed that the smart city network management of urban governance faces the dilemma of technology, thinking, organization and value [31], and should seek optimization from top-level system design, urban organization and structure, people's subjectivity, and human-technology cooperation division of labor [32, 33]. To sum up, the domestic research on the smart city network management city has entered a preliminary mature stage in the field of local design and development and application, and the exploration of the smart city network management city construction mechanism and governance ideas needs to be further increased and improved.

3 RESEARCH ON FUSION PROCESSING AND ANALYSIS ALGORITHM OF HETEROGENEOUS DATA FROM MULTIPLE SOURCES

Urban computing takes data mining and artificial intelligence as technical means to sense, manage and analyze multi-source heterogeneous data, so as to solve the problems and challenges in the process of rapid urban development. The overall framework of urban computing is shown in Fig. 1.

City time series data: Time series of attribute observations in a city, such as historical air quality index. The observation of multiple attributes of a city will produce multiple time series. The difficulty of urban time series data fusion lies in how to effectively integrate time series data with different trends. In this paper, a data fusion method based on semantic mapping is designed to integrate multiple time series data.

Urban spatial data: A collection of urban Spaces that are observed on each timestamp or in a short period of time without changing over time, such as a Point of Interest POI. The difficulty of urban spatial data fusion lies in how to dig out the correlation between spatial units through different attributes of urban spatial units. In this paper, a new data fusion method based on deep reinforcement learning is proposed to realize the fusion of multiple spatial data.

In this paper, a new cross-domain knowledge graph cross-embedding method is designed to integrate the knowledge graph of multiple fields in the city.

Urban time-space data and knowledge graph: By integrating time series data, spatial data and knowledge graph, urban spatial-temporal information stored in the form of knowledge triples is realized, such as the historical traffic changes between any two regions stored in the form of knowledge triples. The existing urban knowledge maps generally use natural language processing technology to extract knowledge from text data, but could not dig out spatio-temporal knowledge. Based on deep learning and knowledge graph technology, this paper constructs a fusion model of urban time-space data and knowledge graph, and realizes the collaborative fusion of urban temporal data, spatial data and knowledge graph by studying the common representation of multi-source heterogeneous data.

3.1 Multi-Source Heterogeneous Urban Spatial Data Fusion Algorithm of Ontology

The entity alignment model based on bidirectional GCN (Graph Convolutional Networks) and CVM (Cloud Virtual Machine) is adopted to mine entity semantic information by fully considering the relationship between entities. As shown in Fig. 2, the model framework consists of three modules, namely, surface structure embedding module, deep semantic embedding module and entity alignment module. First, in the surface structure embedding module, the knowledge graph is divided into entities as subgraphs in two states, and then put into GCN to learn entity embedding, and the global features of entity structure are synthesized. Secondly, in the deep semantic embedding module, we use CVM to weight important attributes and aggregate the attribute values under different weights to get the feature embedding of the entity semantic layer. Finally, in the entity alignment module, the random gradient descent method is used to train the module to determine whether the two entities are aligned, as shown in Fig. 2.

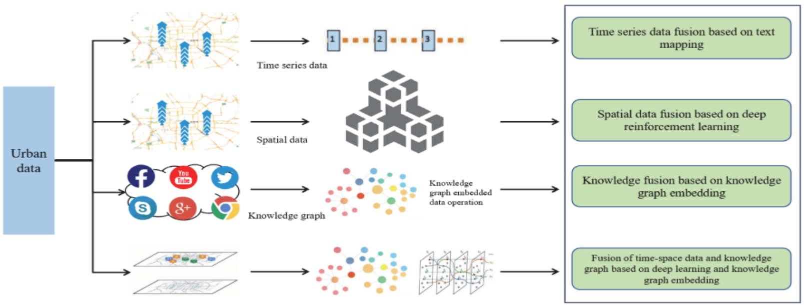


Figure 1 Research framework of multi-source heterogeneous city data fusion technology

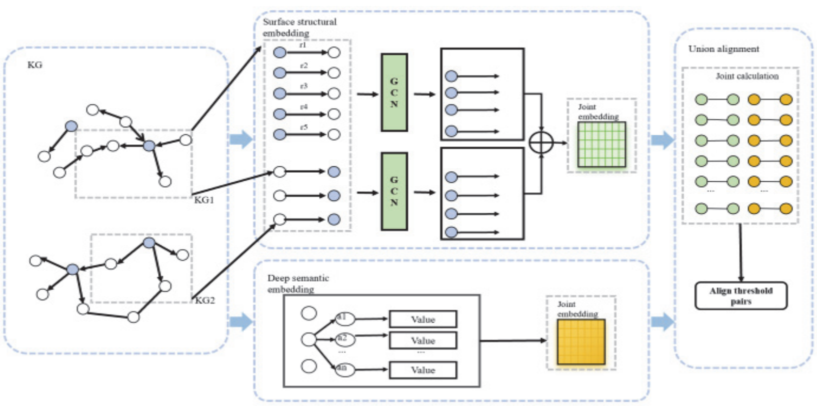


Figure 2 Block diagram of entity alignment model

Encode the words in the vocabulary 0 - 1, denoting them as x, and then use the mean logarithmic conditional probability *Pt* to maximize the word embedding representation, as follows:

$$Pt = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{c} \lg p(x_{t+j} \mid x_t) \tag{1}$$

where, c is the size of the training sample window; X_{t-j} and X_{t+j} are the central concepts. There are j concepts before and after X_t. T is the total number of concepts in the training sentence. Use the SoftmaX function to define the probability function as:

$$p(x_{t+j} \mid x_t) = \frac{\exp(y_{t+j}, y_{x_t})}{\sum_{x=1}^{t} \exp(y_x, y_{x_t})} \tag{2}$$

For N entities, this paper adopts 300 dimensional [30] entity features to form $N \times 300$ dimensional feature matrix X, and N entities to form $N \times N$ dimensional neighbor weight matrix A. Relation: The relation information from the first node to the JTH node in the adjacent-weight matrix. The adjacent-weight matrix defines the neighbors of entities in the convolution calculation. In the heterogeneous knowledge graph, there are different directions and types of the corresponding entity association relations. This paper distinguishes the influence degree of different relations on entities by weighting the relations. The relationship weight is defined as:

$$w_r = \max\left({n_r}/{n_R}, r \right) \tag{3}$$

where n_R is the total number of relation types, n_r is the number of occurrences of a certain relation r, and r is the hyperparameter used to prevent the error problem caused by matrix imbalance caused by too large a difference in weight values.

For each entity i in KG1 and each entity j in K2, the distance function of the joint structure and attribute is:

$$D(x_i, x_j) = rf(x_i, x_j) + (1-r) f(x_{tx}, x_{tj}) \tag{4}$$

$f(x, y)$ is the calculation of entity embedding similarity, or is the entity structure layer embedding representation, et is the semantic layer embedding representation, and B is the hyperparameter that balances the importance of the two types of embedding. We choose the method of negative entity pairs and find the negative entity pairs which are slightly different from the positive example to train the model in this paper, which makes the training more challenging. The minimization margin loss function is adopted for model training, which is defined as follows:

$$L = \sum_{(x_i, x_j)} \sum_{(x_{ti}, x_{tj})} D(x_i, x_j) + r * D(x_{ti}, x_{tj}) \tag{5}$$

In this paper, the random gradient descent (SGD) method is chosen to minimize the loss function, and the distance between the positive and negative pairs is as close as possible, so as to exclude the dissimilar entities.

3.2 Smart City Multi-Source Heterogeneous Spatio-Temporal Data Fusion Processing and Analysis

Urban multi-source heterogeneous data is shown in Fig. 3. Deep learning output cannot be effectively interpreted, which is also a barrier to the fusion of heterogeneous data from multiple sources. In this case, how to extract the features of temporal data, spatial data and knowledge triples into the same semantic space is a technical difficulty.

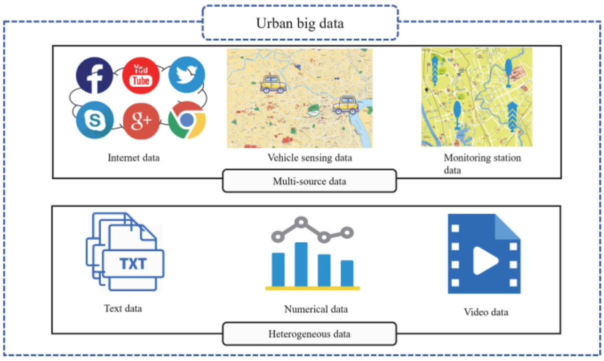


Figure 3 Urban multi-source heterogeneous data

For the fusion of regional traffic patterns and regional POI semantic features, attention mechanism is used to strengthen the POI semantic features of specific two regions, which is called attention-based fusion. The POI semantic feature P of the region is obtained by ID-CNN autoencoder model. In order to fuse it with the regional flow pattern between the two regions, a new P&I Trans layer is first defined to convert it into the characteristics of inter-regional POI changes. Attention-based fusion is mainly divided into two layers, namely attention layer and connection fusion layer. The POI semantic features of the regions are input to the attention layer after passing through the POITrans layer. Given a query Q and a set of key values K, the attention score is calculated by using dot products to calculate non-normalized salient features.

$$Attention = soft\max\left(\frac{QK}{\sqrt{d}}\right) * r \tag{6}$$

where d is the dimension of the key vector as a scale factor. Regional POI semantic features with attention are fused with regional traffic patterns by connecting fusion layers.

In the first module, the goal is to obtain embedded representations of relationships and entities through deep learning models. For LSTM autoencoders, 1D-CNN autoencoders, and two 2D-CNN autoencoders, the mean square error of the input sample (x) and the output value (y) is minimized.

$$Loss = \frac{1}{n}\sum_{i} \| x_i - y_i \|_2^F \tag{7}$$

where *Loss* is the loss function of the LSTM autoencoder and nD-CNN autoencoder (n =1 or 2) models, where n is the number of samples. The || parameter represents the Frobenius norm.

The regional traffic pattern represents the traffic migration trend between two regions in a certain period of time. Regional traffic patterns should be more similar relative to other regions. Therefore, the validity of R in regional traffic pattern knowledge is measured by calculating the similarity of regional traffic patterns in the same two regions on different dates. On the other hand, for entity set V, the differences in the embedding representations of regions at different times are caused by the irreversible relations of regions. Therefore, the similarity of embedded representations of the same region on different days is calculated in the same way to measure the validity of V in the knowledge of regional traffic patterns. Cosine similarity evaluates the similarity of two vectors by calculating the Angle cosine of the two vectors. Therefore, it is used to calculate the similarity of different date entities and the similarity of relationships respectively. The greater the similarity of entities or relationships r, the greater the validity of regional traffic pattern knowledge.

$$sim_R = \frac{1}{|R|} * \frac{r_{ik} r_{jk}}{\| r_{ik} \|_2^F \| r_{jk} \|_2^F} \tag{8}$$

$$sim_v = \frac{1}{|V|} * \frac{v_{ik} v_{jk}}{\| v_{ik} \|_2^F \| v_{jk} \|_2^F} \tag{9}$$

Tab. 1 illustrates the similarity of the relationship between the different days. Where, $days_i$ represents the i day, where $days_{1,3}$ shows the similarity of different working day relationships, and $days_{1,7}$ shows the similarity of weekend relationships.

Table 1 Similarity of relationships between different days

Model	$days_{1,2}$	$days_{1,3}$	$Days_{1,6}$	$Days_{1,7}$	$Days_{1,8}$	daysavg
MF-based	0.6617	0.6434	0.5618	0.5648	0.6729	0.6216
VAE-based	0.7135	0.6770	0 .6941	0.6906	0.7056	0.6961
RFP-KMNnopor	0.7689	0.7794	0.7386	0.7292	0.7645	0.7562
RFP-KMN	0.8614	0.8428	0.8204	0.8129	0.8652	0.8405

The experimental results are analyzed vertically and horizontally in Tab. 1. From the longitudinal analysis in Tab. 1, the similarity of MF-based methods is lower than other methods. This method has some limitations for extracting time series features. What is more, the VAE based approach performs better than the MF-based approach, but worse than the model proposed in this chapter. It can be seen from the similarity of RFP-KMN that regional POI characteristics have a great impact on inter-regional traffic changes. Intuitively, the flow variation of an area is usually related to its POI characteristics. The proposed RFP-KMN model integrates POI features, which makes the embedded representation of relations achieve better results. From the horizontal analysis of Tab. 1, $day_{1,6}$ and $days_{1,7}$ are smaller than other values, which indicates that weekends also have a certain impact on regional traffic changes. However, in practice, only some areas of traffic changes are affected by holidays. The POI characteristics of these regions are similar to those of other regions. Therefore, the variation of regional flow is affected by many factors.

Tab. 2 shows the similarity of entities across days. The experimental results in Tab. 2 were analyzed in the same manner. From the longitudinal analysis of Tab. 2, the similarity between MF-based methods, VAE based methods and RFP-KMNnoRd is low. The reversible feature (Rd) of a region is a stable feature that enhances the embedded representation of the region. It can be seen from the results of RFP-KMN that the POI feature of the region can also enhance the embedded representation of the region. In contrast, POI characteristics have a greater impact on relationships than on entities. This also vaguely reflects that the regional traffic change trend is greatly affected by POI characteristics. From the horizontal analysis in Tab. 2, according to the values of day 1,6 and days 1,7, it can be seen that the weekend has little influence on the embedding representation of the region.

Table 2 Similarity of entities between different days

Model	$days_{1,2}$	$days_{1,3}$	$Days_{1,6}$	$Days_{1,7}$	$Days_{1,8}$	daysavg
MF-based	0.7289	0.8324	0.7122	0.7323	0.7439	0.7583
VAE-based	0.7615	0.9124	0.8035	0.8099	0.8237	0.8416
RFP-KMNnopor	0.8549	0.8315	0.7214	0.7214	0.7813	0.7634
RFP-KMN	0.8729	0.9462	0.8712	0.8762	0.8782	0.8848

In the traffic prediction task, the traffic prediction results of different models are shown in Fig. 4. Where, "-RD "indicates that the input data of the model is the original data (RD), and "-TFP" indicates that the input data of the model is the traffic flow mode (TFP). When predicting traffic from raw data, LGBM predicted the best results, followed by XGBoost.

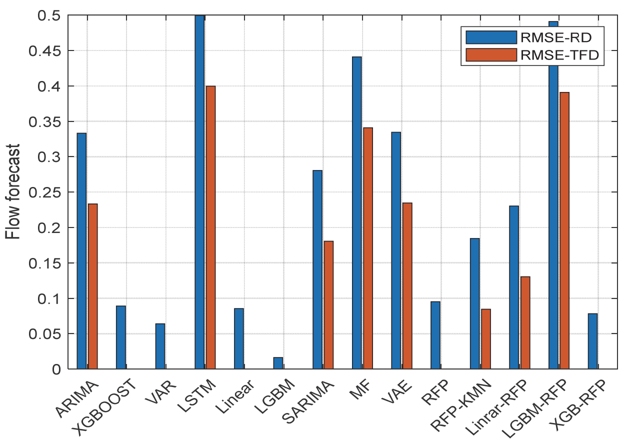


Figure 4 Results of traffic prediction

For different traffic flow pattern mining models, Linear (RFP-KMN) is better than Linear (RFP-KMNnoPOI), indicating that the traffic in different time periods between regions is affected by regional POI characteristics, and regional POI characteristics can improve the traffic prediction results. The results of Linear (MF-based) and Linear (VEV-based) prediction are not as good as those of Linear (RFP-KMN) model, indicating that Linear (RFP-KMN) is more suitable for extracting features of time series data. In addition, the prediction results of LSTM(RFP-KMN) are not as good as those of Linear (RFP-KMN). The possible reason is that the traffic flow pattern has already been trained with a deep learning-based model, and if LSTM is used again, it will lead to overfitting.

4 RESEARCH ON SMART CITY ONE-NETWORK MANAGEMENT AFTER FUSION TREATMENT

4.1 Modernization of Unified Management and Governance

The architecture and positioning of metallurgical modernization are shown in Fig. 5. After the fusion of multi-source heterogeneous data, the smart city one-network management and management enables instructions to be transferred between different departments to complete efficient and collaborative handling of events. At the same time, set up a smart city network management and governance command center entity after the fusion of multi-source heterogeneous data, integrate the urban management, comprehensive management, hotline 12345 and other relevant departments with it, and integrate isolated and dispersed forces into an efficient collaborative team to govern the city through the dual wheel drive of mechanism innovation and technological innovation.

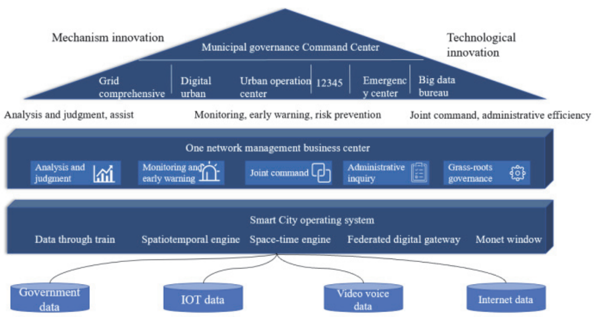


Figure 5 Architecture and positioning of smart city one-network management and governance modernization after multi-source heterogeneous data fusion

After the fusion of multi-source heterogeneous data, the smart city one-network management and governance provides four major values for city managers: analysis, research and judgment to assist decision-making, monitoring and early warning to prevent risks, linkage command administrative efficiency, and grassroots governance to build and share.

1) Combine the industry knowledge of experts to efficiently generate analysis and judgment reports to assist government decision-making. The content of the report includes three parts: quantification of the current situation, cause analysis and strategy suggestions.

2) Build an early warning model to support accurate discovery and timely disposal of early warning events and help the government prevent risks.

3) Administrative efficiency of linkage command: realize linkage command at the city, district/county, street/township and community levels, shorten the decision-making process and response chain, ensure non-destructive information transmission and consistent actions at all levels, and effectively coordinate multi-departments. At the same time, the work efficiency of the department is assessed through the record of incident handling, and the existing incident handling process is further optimized.

4) Co-construction and sharing of grassroots governance: through the one-network management system, the information channel between the government and residents, to achieve mass prevention and treatment, co-construction and sharing, such as residents can report the problems they find to the one-network management system in a timely manner through 12345 and other applications, to help the government to jointly govern the city.

4.2 Practical Value of Constructing a Smart City System After the Integration of Multi-Source Heterogeneous Data With One Network

1. Improve data quality and reliability.

Establish a smart city system after the integration of multi-source heterogeneous data.

2. Support scientific and refined decision-making.

Establish a smart city system after the integration of multi-source heterogeneous data, and the government can analyze and forecast based on real and accurate data.

3. Promote information sharing and collaborative working.

The establishment of a smart city system after the integration of multi-source heterogeneous data, government departments can share data resources, avoid the phenomenon of data "island", and improve work efficiency and synergy.

4. Strengthen data security and privacy protection.

Establish a smart city system after the integration of multi-source heterogeneous data, effectively prevent data leakage and abuse, and protect citizens' personal privacy and data security.

5. Promote open and innovative use of data.

The establishment of a smart city system after the integration of multi-source and heterogeneous data is conducive to the open sharing and innovative application of government data. The government can rely on a unified data platform to open data resources to the society, stimulate the innovative application and value mining of data, and promote the development of digital economy and scientific and technological innovation.

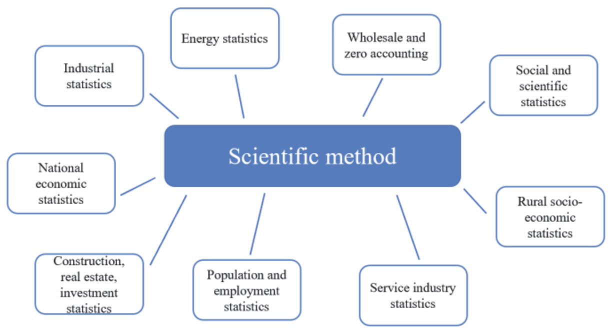


Figure 6 Scientific method

Build a smart city database after the integration of important multi-source heterogeneous data such as "enterprise basic database", "standard four-up enterprise database" and "quasi-four-up enterprise database" based on shared organizational "identification information" and "attribute information", further understand the province's economic "background", and improve the scientific, systematic and complete statistics. We will promote multi-departmental interaction and coordination to do a good job in monitoring, management and statistics of economic operations. The scientific method is shown in Fig. 6.

The evaluation index of smart city system after the integration of multi-source heterogeneous data can verify the advantages and disadvantages of a method. In this experiment, the main work is to align entities in large-scale heterogeneous data sets. The selection of hyperparameters affects the alignment performance of the model. In the experiment, the initial dimension of entity embedding and the alignment dimension are set to 300 dimensions. In the training, the parameters are iteratively updated by stochastic gradient descent until convergence. The ratio parameter B for the combination of surface structure and deep semantics of the experimental entity is set between 0 and 1, Y is set to 1, the learning rate $1r$ is set to 3.0, and the epoch number of iterations is set to 500. The closest entity is selected by Euclidean distance, and a negative example is generated by randomly replacing one of the positive entity pairs.

As the amount of attribute information varies in different data sets, the choice of B value varies accordingly. As can be seen from Fig. 7, the size of B value has different influences on the alignment accuracy of the two groups of data.

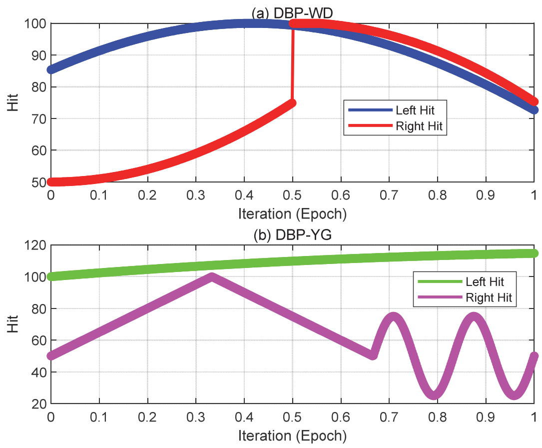


Figure. 7 Changes of evaluation index with p value

5 SIMULATION VERIFICATION

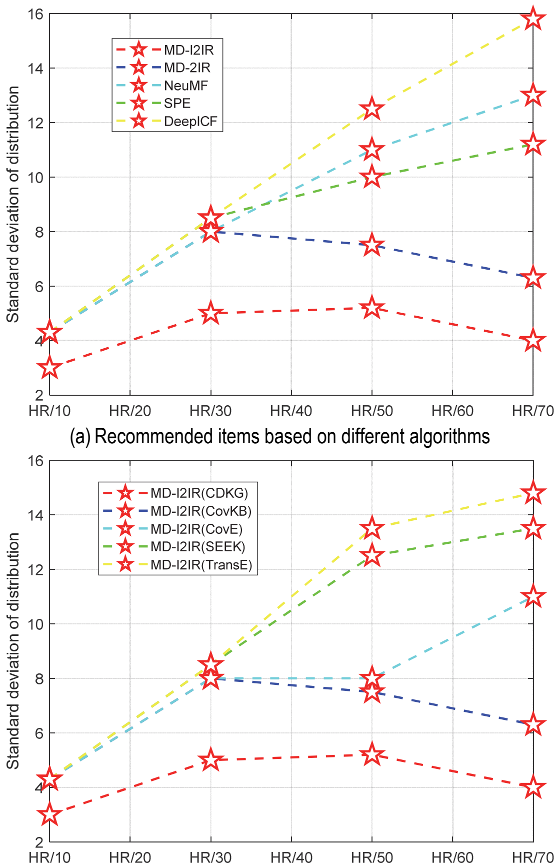
In the "one network management" application, it covers seven basic governance service applications. Take resource integration as the main line, focus on grassroots grid business as the key direction, and combine the hot and difficult issues faced by Hainan Xueliang Project and grassroots governance. Carry out the construction of mass prevention and mass governance application systems (mainly including multi-network co-governance comprehensive management platform, comprehensive governance video research and judgment analysis platform, grassroots public security comprehensive management platform, comprehensive convenience service platform, social autonomy and service management platform, grassroots integrated command platform, micro political and legal platform), and gather all kinds of grassroots video information, business information and early warning information through the government affairs external network. Carry out basic governance work such as grassroots grid data collection, grid organization management, grid business development, social autonomy and grassroots party building of citizens, and mass demands for the seven grassroots innovation applications, and improve the modernization level of grassroots community governance system and governance capacity. Through the construction of this platform, we can play the role of mass prevention and treatment, rely on law enforcement forces such as grassroots staff, grid personnel, inspectors, and social forces such as volunteers and the masses to achieve full video coverage of grassroots communities, improve the level of grassroots governance, and improve the sense of gain, happiness and satisfaction of the people. We will foster a safe political environment, a stable social environment and a quality business environment. Application construction and application need to establish a special database of mass prevention and treatment service based on data resource sharing. Based on the requirements of the "Standard for the Open System of the Data Center of the General Platform of Social Management", the special database of the group prevention and management business is established, which is an important part of the data resource pool of the general platform of social management.

Tab. 3 shows the results of comparison between the model proposed in this paper and other benchmark models. In general, the model proposed in this paper is superior to other methods on NYT24 and NYT29 data sets. The experimental results of the comparison model are derived from the data in the paper, and it can be seen from the comparison model that the results of the WDec model are the best performing of all the baseline models. Compared with the WDec model, the $F1$ value of the model in this paper improves by 2.9% and 10.5%, respectively, on the two versions of the dataset. For the first time, Tagging turns joint extraction into a sequential labeling problem. Compared with tagging results, the proposed method significantly improves the results in all three indexes, because multiple extraction is carried out under each relation, and a feedback mechanism is added to prevent the same entity pairs from being extracted, effectively extracting overlapping triples. The CopyRE model can only extract the last word of the entity, but the recall rate and $F1$ value are improved compared with tagging. In this paper, the pointer network is used to decode the position pointer of the entity in the sentence, and the full name of the entity can be extracted, overcoming this difficulty and achieving better results. The relational attention mechanism is added to the method in this paper, which can overcome the large number of irrelevant entities extracted by HRL and reduce the model's attention to irrelevant entities.

Table 3 Comparison of training results of NYT dataset

Model	NYT24			NYT29		
	Accuracy	Recall	$F1$	Accuracy	Recall	$F1$
Tagging	0.624	0.315	0.422	0.594	0.382	0.464
CopyRE	0.612	0.563	0.586	0.563	0.453	0.505
HRL	0.783	0.772	0.773	0.695	0.603	0.644
PNDec	0.807	0.774	0.783	0.733	0.622	0.674
WDec	0.881	0.765	0.816	0.775	0.608	0.683
Textual model	0.862	0.827	0.845	0.810	0.774	0.788

The variation trend of the standard deviation of the distribution with the increase of the number of recommended terms is also analyzed, as shown in Fig. 8.



(a) Recommended items based on different algorithms

(b) Recommended items based on different knowledge maps

Figure 8 Variation of the standard deviation of the distribution as the number of recommended terms increases

In Fig. 8a, the distribution standard error of NeuMF, SPE, Content+topology KNN, DeepICF, and MD-I2IR(ip) increases as the number of recommendations two increases, indicating that the distribution of their recommended terms is increasingly unbalanced across different domains. The distribution standard deviation of MD-I2IR(ip) and MD-I2IR increases first and then decreases with the increase of recommendation number two, which indicates that the distribution of their recommended terms in different domains can remain stable. In Fig. 8b, the standard deviation of distribution of MD-I2IR (TransE), MD-I2IR (SEEK-1) and MD-I2IR (ConvE) increases with the increase of recommendation number two, which indicates that the distribution of recommended terms in different domains is relatively unstable. MD-I2IR (ConvKB) can keep the standard deviation of the distribution of recommendations steady when the number of recommendations exceeds 20. Compared with other methods, the proposed MD-I2IR(CDKG-CE) has a lower distribution standard deviation with the increase of the recommended number two, and can maintain the stationality of distribution standard deviation.

In the situation analysis of smart city one-network management, Fig. 9 shows the correlation between image feature Fi and PM2.5. The *X*-axis and *Y*-axis are the features extracted from the image. When the *X*-axis and *Y*-axis are the same, they represent the relationship between single feature and PM2.5. Fig. 9 uses different colors and shapes to represent the classification results. Since the photos were taken from PM2.5 monitoring sites, all of the image data carries real PM2.5 labels, and these features are very differentiating in PM2.5 inference.

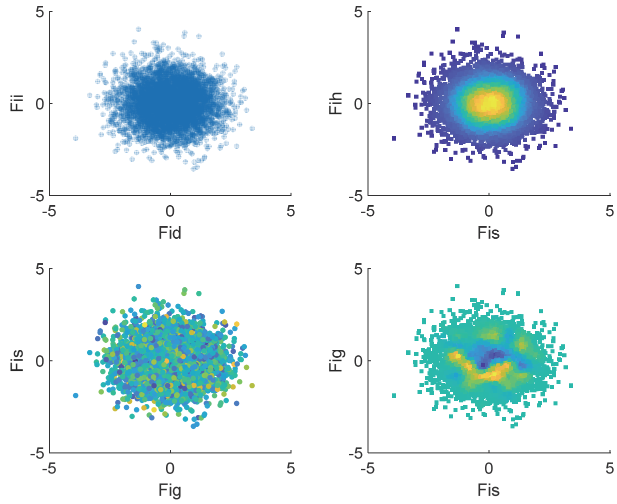


Figure 9 Correlation between Fi and PM2.5

In order to evaluate the effectiveness of the model, it is not enough to consider the overall indicators. Fig. 10 shows the accuracy results, with different models having very different classification effects. The method proposed in this chapter has a great improvement in the classification of good, light pollution and moderate pollution, and has a similar effect as other models in the classification of excellent, severe or severe pollution. The method has significantly improved the mean value. Fig. 10 also shows the results of recall rate. Except that the method of severe or severe contamination is slightly worse than that of naive Bayes, other conclusions are basically the same as the results of accuracy. Fig. 10 also shows the results of comprehensive evaluation indicators, true rate, false positive rate and ROC curve area, and the method has obvious advantages.

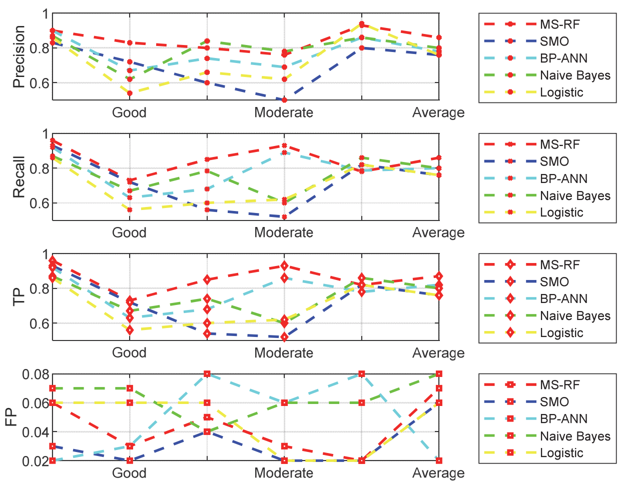


Figure 10 Classification results of various types of weather

Fig. 11 shows the accuracy of classification results of the proposed method. Although the use of image data alone is less effective in assessing air quality, classification accuracy is always significantly improved when image features are added to the model. In particular, image features have a good effect in the classification of moderate pollution, and the overall accuracy of moderate pollution is significantly improved after adding image features. Combining multiple data sources is necessary when the problem at hand is very complex. The results of other evaluation indicators are similar to the results of accuracy. As Fig. 11 shows, although the classification results for good weather were slightly worse, the classification recall rate for moderate pollution increased by 83%.

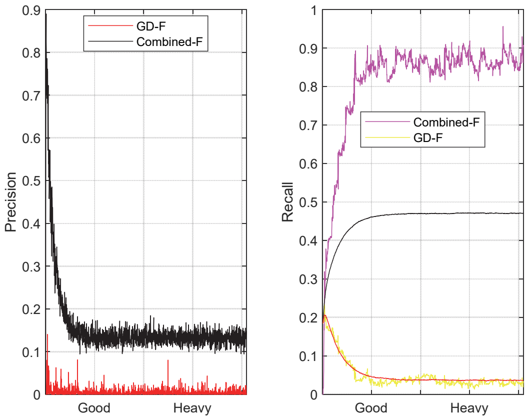


Figure 11 Influence of various features on classification results

As shown in Fig. 11, the false positive rate and the area of the curve, as well as the classification effect of light pollution, moderate pollution and heavy pollution have been significantly improved after comprehensive consideration of multiple data sources.

6 CONCLUSION

In this paper, we design and implement a relational extraction model based on ontology-based pointer network joint decoding. Attention mechanisms are used to decode entities in corpus sentences and to reinforce interactions between entities and relationships. For decoding, two pointer networks are used to extract the header and tail entities respectively under each relation, which reduces the model's attention to unrelated entities, and can extract entity names of different lengths. Finally, experiments on two versions of NYT data sets verify the validity of the model. The entity alignment model of joint structure and attribute is proposed. In the smart city environment, based on the sensor data features of dynamic target recognition, according to the results of machine learning methods to identify dynamic target features, ontology-based feature-level data description is constructed to provide ontology data for the higher-level data fusion stage. This paper also proposes the case updating algorithm of sensor ontology and the attribute fusion and feature extraction algorithm based on ontology instance. In order to evaluate the effectiveness of the multi-source heterogeneous data fusion method based on model integration proposed in this paper, a time classifier, real-time data classifier, spatial classifier and real-time data classifier are constructed based on the meteorological data, air quality data, traffic data, PEI data and image data of the participatory perception platform. Furthermore, the neural network based on ELM extreme learning machine is used for model integration, which achieves higher evaluation accuracy than the multi-source heterogeneous data fusion method based on random forest and other existing research schemes proposed in this paper. The next step will be to fully empower the layer, create a data center and a business center. The data center will take the big data resource platform as the core, build a unified support platform around common technologies such as the Internet of Things, video, geographic information, and artificial intelligence in the operation of the park, drive the formation of data services and data intelligence, and provide continuous data-based empowerment for upper-layer applications.

Acknowledgments

This research was supported by the Hainan Provincial government affairs collaborative office platform (NO.59901320-3, 59900422, 59900623).

7 REFERENCES

[1] Chen, Y., Wang, C., & Zhou, Y. (2024). Research on multi-source heterogeneous big data fusion method based on feature level. *International Journal of Pattern Recognition & Artificial Intelligence, 38*(2), 12-34. https://doi.org/10.1142/S0218001424550012

[2] Wan, Y., Li, D., & Li, Z. (2024). A semi-supervised four-chamber echocardiographic video segmentation algorithm based on multilevel edge perception and calibration fusion. *Ultrasound in Medicine & Biology, 50*(9), 1308-1317. https://doi.org/10.1016/j.ultrasmedbio.2024.04.013

[3] Yuan, Y., Liu, X., & Lu, K. (2024). Multi-perspective data fusion framework based on hierarchical BERT: Provide visual predictions of business processes. *Computers, Materials & Continua, 78*(1), 46937-46958. https://doi.org/10.32604/cmc.2023.046937

[4] Huangfu, P. & Dang, L. (2023). A multi-scale pyramid feature fusion-based object detection method for remote sensing images. *International Journal of Remote Sensing, 44*(23-24), 7790-7807. https://doi.org/10.1080/01431161.2023.2288947

[5] Lu, S. (2021). Research on computer programming optimization system based on big data technology. *Journal of Physics: Conference Series, 1802*(3), 32046-32053. https://doi.org/10.1088/1742-6596/1802/3/032046

[6] Chen, Z., Deng, L., & Gou, J.(2023).Building and road detection from remote sensing images based on weights adaptive multi-teacher collaborative distillation using a fused knowledge. *International Journal of Applied Earth Observation and Geoinformation, 124*, 3522-3545. https://doi.org/10.1016/j.jag.2023.103522.

[7] Weina, Z., Mingrun, Z., & Qihao, S. W. (2023). Integrated analysis of energy carbon emissions and air pollution in Ningxia based on MGWR and multisource remote sensing data. *Arabian journal of geosciences, 16*(9), 616-622. https://doi.org/10.1007/s12517-023-11616-6

[8] Xi, C., Kaoru, H., & Zhiyang, Y. J. (2023). A model fusion method based on multi-source heterogeneous data for stock trading signal prediction. *Soft Computing-A Fusion of Foundations, Methodologies & Applications, 27*(10), 7714-7718. https://doi.org/10.1007/s00500-022-07714-4

[9] Hong, R.(2024).Analysis of Factors Affecting the Accuracy of MFD Construction in Multisource Complex Data Scenarios. *Sustainability, 16*, 8018-8032. https://doi.org/10.3390/su16188018

[10] Kamm, S., Veekati, S. S., & Timo, M. (2023).A survey on machine learning based analysis of heterogeneous data in industrial automation. *Comput. Ind. 149*, 103930-103952. https://doi.org/10.1016/j.compind.2023.103930

[11] Xiao, Y., Zhang, J., & Chi, C.(2023).Criticality and clinical department prediction of ED patients using machine learning based on heterogeneous medical data. *Computers in Biology and Medicine, 165*, 7390-7404. https://doi.org/10.1016/j.compbiomed.2023.107390

[12] Dasappa, N. S., Kumar, G. K., & Somu, N. (2024). Multi-sensor data fusion framework for energy optimization in smart homes. *Renewable & sustainable energy reviews*, *193*(Apr.), 1.1-1.14. https://doi.org/10.1016/j.rser.2023.114235

[13] Chen, X., Hirota, K., & Dai, Y. (2022). A model fusion method based on multi-source heterogeneous data for stock trading signal prediction. *Soft Computing, 27*(10), 6587-6611. https://doi.org/10.1007/s00500-022-07714-4

[14] Zhu, Y., Zuo, Y., & Li, T. (2021). Modeling of Ship Fuel Consumption Based on Multisource and Heterogeneous Data: Case Study of Passenger Ship. *Journal of Marine Science and Engineering, 2021*(3), 273-298. https://doi.org/10.3390/JMSE9030273

[15] Liu, J., Li, T., & Ji, S. (2021). Urban flow pattern mining based on multi-source heterogeneous data fusion and knowledge graph embedding. *IEEE Transactions on Knowledge and Data Engineering*, 1041-4347. https://doi.org/10.1109/TKDE.2021.3098612

[16] Tang, M. C., Cao, J., & Gong, D. Q. (2024). Unsupervised Learning-Based Exploration of Urban Rail Transit Passenger Flow Characteristics and Travel Pattern Mining. *International Journal of Computers, Communications & Control*, *19*(2), 6422-6435. https://doi.org/10.15837/ijccc.2024.2.6422

[17] Zhang, Y., Zheng, X., & Chen, M. (2021). Urban Fine-Grained Spatial Structure Detection Based on a New Traffic Flow Interaction Analysis Framework. *International Journal of Geo-Information, 10*(4), 227-245. https://doi.org/10.3390/ijgi10040227

[18] Guo, X., Xu, Z., & Zhang, J. (2020). An OD Flow Clustering Method Based on Vector Constraints: A Case Study for Beijing Taxi Origin-Destination Data. *International Journal of Geo-Information, 9*(2), 128-143. https://doi.org/10.3390/ijgi9020128

[19] Yang, H., Li, S., & Peng, D.(2024). Modeling and Analysis of OFDMA-NOMA-RA Protocol Considering Imperfect SIC in Multi-User Uplink WLANs. *Computers, Materials & Continua, 79*(6), 5273-5294. https://doi.org/10.32604/cmc.2024.050869

[20] Wu, J. (2024). Research on optimization of e-commerce supply chain logistics service model based on multi-source data fusion. *Applied Mathematics and Nonlinear Science*s, 9(1), 1619-1634. https://doi.org/10.2478/amns-2024-1619

[21] Xu, S., Chen, J., & Wu, M. (2021). E-Commerce Supply Chain Process Optimization Based on Whole-Process Sharing of Internet of Things Identification Technology. *Computer Modeling in Engineering and Sciences, 126*(2), 843-854. https://doi.org/10.32604/cmes.2021.014265.

[22] Madhubabu, K. & Snehalatha, N. (2023). Congestion avoidance for electrically charged autonomous vehicles in vehicular Ad hoc network. *International journal of systems assurance engineering and management, 14*(6), 2447-2459. https://doi.org/10.1109/TVT.2017.2755504

[23] Pan, R., Huang, Y., & Xiao, X. (2021). Evaluating Consumers' Willingness to Pay for Delay Compensation Services in Intra-City Delivery-A Value Optimization Study Using Choice. *Information (Switzerland), 12*(3), 127-143. https://doi.org/10.3390/info12030127

[24] Zhang, W. & Tuo, K. (2023). Research on Offloading Strategy for Mobile Edge Computing Based on Improved Grey Wolf Optimization Algorithm. *Electronics (2079-9292), 12*(11), 2533-2549. https://doi.org/10.3390/electronics12112533

[25] Rahul, S. & Bhardwaj, V. (2024). (EERO) Energy-Efficient Fog Resource Optimization Model for Scientific Workflow Applications. *International Journal of Engineering Trends and Technology, 72*(5), 149-164. https://doi.org/10.1109/IROS.2024.281971

[26] Babaei, S. & Hamidieh, A. (2022). Designing a Resilient Location-Allocation Model for Cell Site Networks with Regional Coverage Enhancement Approach Using Robust Programming-Lagrangian Relaxation. *Journal of Applied Research on Industrial Engineering, 9*(4), 2810-2823.

[27] Li, H., Xu, Y., & Peng, J. S. J. (2024). Research on Transparent Access Technology of Government Big Data. *Technical Gazette/Tehnički Vjesnik, 31*(3), 775-791. https://doi.org/10.17559/TV-20230630000775

[28] Yang, B. & Yang, M. (2022). Research on enterprise knowledge service based on semantic reasoning and data fusion. *Neural computing & applications, 34*(12), 9455-9470. https://doi.org/10.1007/s00521-021-06382-z

[29] Meng, J. X., Ding, G., & Liu, L. (2021). Research on a Prediction Method for Carbon Dioxide Concentration Based on an Optimized LSTM Network of Spatio-Temporal Data Fusion. *IEICE Trans. Inf. Syst., 104-D*, 1753-1757. https://doi.org/10.1587/transinf.2021edl8020

[30] Zhao, R., Wang, Y., & Liu, Q. (2020). Analysis of crowd flow characteristics based on multi-space scale and multi-source data fusion. *IOP Conference Series Materials Science and Engineering, 768*, 52027:52048. https://doi.org/10.1088/1757-899X/768/5/052027

[31] Strelet, E., Peng, Y., & Castillo, I. (2023). Multi-source and multimodal data fusion for improved management of a wastewater treatment plant. *Journal of Environmental Chemical Engineering, 11*(6), 530-547. https://doi.org/10.1016/j.jece.2023.111530

[32] Fan, Y., Lv, Z., Chen, S., Yao, C., Yuxin, W., Xuan, Z., Zhiqiang, X., & Weidong, W. (2025). Condition recognition based on multi-source heterogeneous data and residual temporal network in coal flotation process. *IOP Publishing Ltd*. https://doi.org/10.1088/1361-6501/adb064

[33] Zhou, K., Lu, N., Jiang, B., & Zhisheng, Y. (2025). FEV-Swin: Multi-source heterogeneous information fusion under a variant swin transformer framework for intelligent cross-domain fault diagnosis. *Knowledge-Based Systems*, *310*. https://doi.org/10.1016/j.knosys.2025.112982

Contact information:

Hang LI
Big Data Development Center of Hainan Province,
Haikou, 570204, China

Xiao XIAO
(Corresponding author)
School of Philosophy, Wuhan University,
Wuhan 430072, China
E-mail: xiaoxiao_whu@163.com