International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 1087-1096

https://doi.org/10.70517/ijhsa464101

Research on the Path of Administrative Efficiency Improvement under the Construction of Digital Government

Naiqing Bu[1,*]
[1] School of Social Sciences Sanya College, Sanya, Hainan, 572000, China
Corresponding authors: (e-mail: buqing2024@126.com).

Abstract In the transformation of government functions to “service-oriented” construction, administrative efficiency is always the focus of the digital government project. This paper takes the collection, mining and integration of event data as the path to improve administrative efficiency under the digital government platform. The government event collection channels are sorted out and an event distribution model is formed to build a digital platform for government event distribution. Subsequently, the key technology framework for multi-source heterogeneous data fusion of sensitive data is established by cleaning inferior data and standardizing storage of integrated data. For the multi-source heterogeneous data provided by the technical framework, a heterogeneous database log parsing algorithm is proposed to meet the demand for change log data capture from multiple heterogeneous databases for the daily operation of government affairs. After completing the data preparation, a multi-source heterogeneous data mining model is constructed to carry out multi-source heterogeneous data mining based on fuzzy C-mean clustering to realize the deep mining of multi-source heterogeneous few class data sets. Compared with similar model algorithms, the data screening time of the multi-source heterogeneous data mining model is always under 20s, which assists in improving the administrative efficiency of the digital government platform with superior data processing speed.

Index Terms digital government, multi-source heterogeneous data mining, database log parsing, administrative effectiveness, event allocation modeling

I. Introduction

Administrative process is a collection of interrelated activities carried out by the government or other authorized authorities to deal with a certain government service [1]. In order to further improve the efficiency of the government in the process of administrative processes, and at the same time to improve and perfect the public service system, creating a digital government has become an important measure and means in the process of perfecting the governmental governance system in countries and regions all over the world [2]-[4].

Digital government has been regarded as an observational indicator to enhance the modernization of government performance and governance capacity, and is an important dimension of the modernization of national governance system and governance capacity [5], [6]. It conforms to the trend of digital change, builds a people-centered digital government operation system, and has remarkable performance in improving decision-making efficiency and reducing administrative costs, which is endogenous to the enhancement of duty performance and externalized to the enhancement of administrative efficiency [7]-[10]. Therefore, the construction of digital government system can have higher superiority in such aspects as more open management system, more efficient collaborative governance, and more effective government operation [11], [12]. Further to establish a government governance system based on scientific and technological means, the construction of digital government has an equally important role to play in realizing the transformation to a service-oriented government [13], [14]. At present, China in the process of digital government construction, has been gradually realized to the public service as the core of the governance model for transformation, digital government has become an important means of building a modern government. In the future, the construction of China's digital government will be centered on improving the administrative efficiency of the government for continuous improvement and development.

In this paper, we first sort out the basic business process of government affairs allocation, establish the event allocation model, and integrate and analyze the event collection channels to propose the event allocation digital platform for government affairs. Secondly, it sequentially describes the multi-source data collection method, the inferior data cleaning process, the standardized storage of integrated data, and the data fusion process, and builds the key technology framework of multi-source heterogeneous data fusion of sensitive data for digital government.

The structure and mathematical implementation methods of log text classifier and context similarity comparator in heterogeneous database are explained again, and the construction process of multi-source heterogeneous data mining model is analyzed. Finally, the operational performance of the designed model is examined and evaluated by comparing similar model algorithms.

II. Construction of a digital platform for the distribution of government incidents

II. A.Presentation of the event allocation model

The event distribution model is the core and key to the construction of the unified event distribution platform, and the complexity of the model has a direct impact on the number of business systems to be integrated and the difficulty of platform construction. The simpler the model is, the fewer systems can be integrated, but the difficulty of the system construction is also smaller, and the stability of the system is higher. On the contrary, the more complex the model is, the more systems can be integrated, but the more difficult the system construction is, and the less stable the system is. Therefore, the establishment of the event distribution model should fully balance the above factors and combine with the business needs to establish a reasonable and feasible event distribution model. Based on the theoretical research results of the four communication types of government hotlines (Q&A, consultative, coordinating and pivotal) in the transformation of urban governance, combined with the sorting out of the communication types of government departments, there are also the above four communication types in the unified event distribution business. Based on the decomposition of communication types and business types, the basic business process of the event distribution model is shown in Figure 1, the event collection and acceptance is transferred to the event distribution link, the distributor will distribute the event to the corresponding responsible unit, the responsible unit to confirm the event, such as not within the scope of responsibility of the unit can choose to return the event, at the same time, the distributor can also be returned to the case of the multiple times in conjunction with the competent department for coordination of the difficult events, clear the responsible unit to the responsible unit, and then to the responsible unit for coordination. After the responsible unit is clarified, it is handed over to the responsible unit for processing, and the responsible unit sends the disposal result notification to the relevant personnel after the disposal of the event is completed, and at the same time closes the case for the event. The basic business process as the main line of the event distribution model runs through the whole process of the distribution and disposal of various types of source events, and the specific business integration process needs to be combined with the different source events, and then personalized design and improvement of the basic business process.

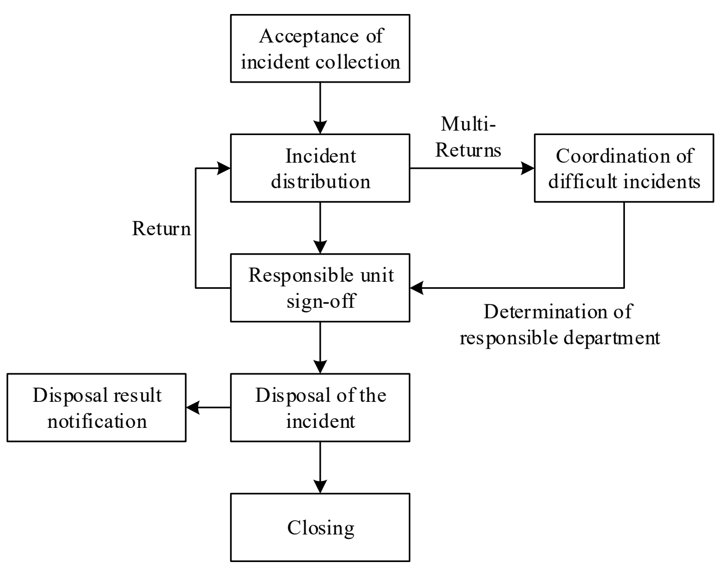

Figure 1: The basic business process of the event allocation model

The event distribution model is based on the idea of integrating government information systems with capability service, encapsulating the work content of departmental business work with common collection, assignment, processing and feedback into the event distribution model, transforming the distribution model into the basic software capability of the information system through information technology, and providing the relevant capability services to the outside world in the form of services through business interfaces, so that users do not need to consider the construction and operation and maintenance problems of the basic capability platform. Users do not need to consider the construction and operation and maintenance of the basic capability platform.

II. B.Integration analysis of event collection channels

The integration of incident collection source channels is the basis for the business integration of the unified incident distribution platform. The integration of source channels requires detailed process combing and analysis of each event distribution business, based on the basic business process of the event distribution model, additional functions are improved, but the adjustment of the main business process needs to be careful. The type of event collection source channel is mainly divided into two types: system docking events and collection and entry events, but no matter what type of event, after entering the unified event distribution platform, the event acceptance needs to be unified coding, definition of event labels, and after completing the above steps, the validity of the event needs to be determined, and for invalid events, they will be directly invalidated or retreated back to the source system or the collection side, and valid events will go to the next link for the validity of the event, and then enter into the next link to carry out the event. For invalid events, they will be invalidated or returned to the source system or the collection side, while valid events will enter the next link, and after the event acceptance, the basic business process of the event distribution model is shown in Figure 2.

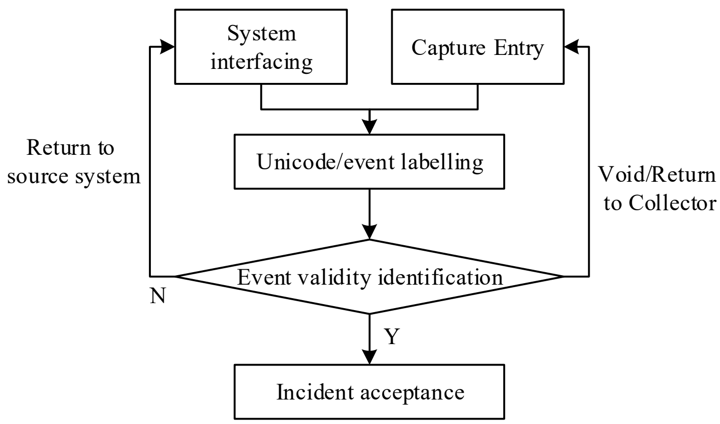

Figure 2: Integrate the sources of event collection and business processes

III. Multi-source Heterogeneous Data Acquisition and Mining

III. A. Key Technical Framework for Multi-source Heterogeneous Data Fusion for Sensitive Data

In recent years, multi-source heterogeneous data fusion has increasingly become a hot issue that has attracted the attention of researchers. Based on the key technologies of multi-source heterogeneous big data processing, this paper studies the key technologies such as defining data structuring, data cleaning, data integration standardization and data fusion, and proposes a set of key technical frameworks for multi-source heterogeneous data fusion including the whole process of "data collection, data cleaning, data integration, and data fusion".

III. A. 1) Multi-source data acquisition

The data sources of governmental big data generally come from departments and bureaus, subordinate business units, governmental exchange platforms or social channels, etc., and the data are collected by means of interface calls from governmental systems, deployment of front-end machines or direct access to data files. In this process, it is very important to ensure the authenticity and integrity of the collected data. Common data sources include: Oracle, MySQL, DB2 and other mainstream databases, Webservice, Kafka, MQ and other types of service interfaces. Forms, documents, pictures and other types of format files.

In the process of data collection, the authenticity and reliability of the collection tool, whether the collector can truthfully and correctly collect the data, the security and authenticity of the front machine, and the security of the repository will affect the authenticity and integrity of the collected data, and there are potential security risks. To do a good job of data collection, both management and technology should be used, not only to formulate corresponding management specifications to provide for the safety management of collection equipment, collection personnel and collection data, but also to automate certain technical tools to ensure the authenticity and completeness of the collected data and to ensure that the collected data are not leaked.

In the data collection scenario, due to the characteristics of big data volume, variety and complexity of sources, it is difficult to verify the authenticity and integrity of the data. At present, there is no strict means of data authenticity, credibility identification and monitoring, and there is the possibility of false or even malicious data in the data, so it is necessary to carry out security protection for the authenticity of the data source and the data collection process. The risk points of governmental big data in the collection process mainly include the data collection tools used, data collection engineers, and data business predecessors.

III. A. 2) Cleaning of poor quality data

A large amount of multi-source heterogeneous data from various business sectors floods the entire network, and cleaning the data, summarizing the structure, usage, and domain of the data, and grading and classifying the data are the basic aspects of data governance. In this paper, firstly, the collected data are cleaned by detecting errors, filling missing and filtering duplicates to ensure the data quality, including the pre-processing process such as checking the data consistency and dealing with invalid and missing values. Secondly, the data is hierarchically classified, and for threatening sensitive data is categorized as private data collection, using encryption technology to set private data access. For data with no security threats, they are categorized as public data sets, and the data sharing scope is set using common access patterns.

III. A. 3) Standardized storage of integration data

The integrated data approach accesses data through an intermediary agent model that provides a unified query structure for the user. The mapping between the mapping agent and the raw data is equivalent to querying each data source for the user. First of all, the data “packaging”, that is, the results of the local query into an easy to handle form, the user in the query, the system will be the user's local query into the corresponding data source query and return the results. Generally speaking, when the data size is large, more frequent updates are more suitable for schema integration methods.

III. A. 4) Data fusion process

Data fusion is the generation of new data through the comprehensive processing of multiple data sets, which will be accompanied by the enhancement or loss of information. Nowadays, it is widely believed that the most commonly used technique for fusion is to use database integration methods to fuse multiple data in the database, while using integrated learning methods to integrate the fused data will obtain data fusion results with high integration and strong generalization ability. The process of data fusion can be specific can collect the attributes of multi-source data for matching the attribute correspondence between the data, i.e., by checking whether the attribute name, attribute type, and attribute value are consistent. On the basis of pattern matching, data referring to the same attribute are connected. Data features are extracted to eliminate the problem of attribute conflict between different data sources. Finally, data from different sources are synthesized and associated at multiple levels through entity fusion, so as to extract more unified, rich and precise data.

III. B. Classifier Structure and Algorithm Implementation

In the implementation part of the algorithm, it is mainly divided into two parts: the implementation of log text classifier and the implementation of context similarity comparator. In the implementation of the classifier, considering the uniqueness of log text, character-level word embedding encoding is used instead of the traditional word embedding method. In the implementation of the comparator, contrast learning is used to optimize the training of the context encoder.

III. B. 1) Log Text Classifier Implementation

The log text classifier consists of an input layer, a hidden layer and an output layer, and the implementation details and improvement aspects of these three layers are explained next.

First, the input layer is responsible for encoding processing and training of word vectors for each word in the original log text separately. Before classifying the log words, special characters (space, comma, period, tab, etc.) are used to segment the original log records. Next, the character-level word embedding model chars2vec, which is the same as the preprocessing part, is used to encode each split word into a word vector form that the hidden layer can receive and process. Compared with the traditional word embedding model, chars2vec is more suitable for log texts with diverse contents and can avoid the problem of vocabulary overflow.

Then, the hidden layer receives the encoded vector after word embedding of the context of the target log word by the input layer, and completes the contextual semantic encoding of the target word by extracting the public templates and semantic features implied in the context through BiGRU. Where h_t denotes the context-hidden state of the target word, and e_t denotes the encoded vector of word embedding of the target log word at the t

moment. Words that are far away from each other in a log record tend to have low correlation with each other and limited contribution to each other's categorization, therefore, the first n and the next n log words of the target word are chosen as the range of the context in the actual training.

The specific procedure of BiGRU for hidden state computation is shown in Equation (1), where R_t is the reset gate and Z_t is the update gate. H_t denotes the hidden state at the current t moment, and $\tilde{H}_t$ denotes the candidate hidden state. ϕ is the activation function that maps the values of Z_t and R_t between 0 and 1. X_t refers to the input vector at moment t, and W_r, W_z, W_h, U_r, U_z, and U_h are the corresponding parameter matrices and cyclic weight matrices for R_t, Z_t, and H_t, respectively. b_r, b_z, and b_h are the corresponding bias vectors for R_t, Z_t, and $\tilde{H}_t$.

$$\begin{aligned} R_t &= \phi\left(X_t W_r + H_{t-1} U_r + b_r\right) \\ Z_t &= \phi\left(X_t W_z + H_{t-1} U_z + b_z\right) \\ \tilde{H}_t &= \tanh\left(X_t W_h + \left(R_t \odot H_{t-1}\right) U_h + b_h\right) \\ H_t &= Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t \end{aligned} \tag{1}$$

Subsequently, the hidden layer inputs the state matrix H_t output from BiGRU into the Attention model to calculate the attention weight assignment. The specific calculation of the attention value is shown in Equation (2). Where e_t is the attention value between each time step, which is calculated from the three attention-related parameter matrices U_t, W_t and b_t with the hidden state matrix H_t at moment t. α_t denotes the attention weight value of the hidden state h_t at moment t, and after calculating the magnitude of the weights of the hidden state at each moment, the contextual final encoding vector s of the current log text is obtained by weighting and summing according to the attention weights.

$$\begin{aligned} e_t &= U_t^T \tanh\left(W_t H_t + b_t\right) \\ \alpha_t &= \frac{\exp(e_t)}{\sum_{k=1}^{n} \exp(e_k)} \\ s &= \sum_{t=1}^{n} \alpha_t h_t \end{aligned} \tag{2}$$

Finally, the output layer is responsible for classifying the encoding results, judging whether the words belong to templates or parameters based on the features of the target log words themselves and their contexts. The output of the hidden layer is input into the sigmoid function to calculate the result of the classification, as in Equation (3), for the classification result.

$$y_i = sigmoid(w_i s_i + b_i) \tag{3}$$

where w_i is the matrix of weight coefficients to be trained from the Attention layer to the output layer, s_i is the output vector of the hidden layer to be classified, b_i denotes the bias to be trained, and y_i is the output classification result, where $y_i = 0$ means that s_i belongs to the type of template, and $y_i = 1$ means that s_i belongs to the type of parameter.

III. B. 2) Implementation of the context similarity comparator

The context similarity comparator improves classification accuracy by assisting in the training of the context encoder in the classifier, with the goal of ensuring that the context encoder generates closer encoding vectors for log records of the same template type, while enabling log records of different template types to generate encoding vectors that are farther apart. The input data to the comparator consists of target log records as query samples and positive and negative samples selected from the results of data preprocessing, and the output is a comparison loss computed based on the similarity between the samples. The comparator uses the back propagation algorithm to update the relevant parameters of the context encoder in the classifier to optimize the model.

The process of calculating the contrast loss of the context similarity comparator: first, calculate the cosine distance between the target log records and the centers of mass of each class cluster in the data preprocessing clustering results, and select the center of mass with the closest distance as a positive sample, and the centers of mass of other class clusters as negative samples. The selected positive and negative samples and query samples are input into the context encoder to generate the corresponding encoding vectors. Then, the cosine similarity between the query sample and the positive and negative samples is computed. Finally, the contrast loss is computed using the Information Noise Contrast Estimation (InfoNCE) loss function. The cosine similarity between two vectors is calculated as shown in Equation (4).

$$\cos_{sim}(A,B) = \frac{A \cdot B}{\|A\|\|B\|} \tag{4}$$

The InfoNCE loss is calculated as in equation (5), where h_q is the context encoding vector of the query sample, h_+ and h_- denote the context encoding vectors of the positive and negative samples, respectively, $k-1$ is the number of negative samples, and the $\cos_{sim}$ function denotes the function that calculates the cosine similarity. It can be seen that the $L_{InfoNCE}$ loss is smaller when the query samples are more similar to the positive samples and more different from the negative samples.

$$L_{InfoNCE} = -\log \frac{\exp(\cos_{sim}(h_q, h_+))}{\exp(\cos_{sim}(h_q, h_+)) + \sum_{i=1}^{k-1} \exp(\cos_{sim}(h_q, h_-))} \tag{5}$$

III. C. Construction of Multi-source Heterogeneous Data Mining Models

In multi-source heterogeneous data, due to the complex structure and overlapping boundaries of data from different sources, it is often difficult to apply traditional clustering methods to deal with them effectively, while fuzzy C-mean clustering (FCM) can deal with multi-source heterogeneous data by assigning a degree of affiliation to each data point to flexibly capture these complex structures and overlapping regions. Therefore, this time, fuzzy C-mean clustering is chosen to realize multi-source heterogeneous data mining in the expectation that more valuable information can be mined from it. Let the dataset after interpolation processing of multi-source heterogeneous data be represented by equation (6):

$$D_{Zi} = \{d_1, d_2, \cdots, d_o\} \tag{6}$$

where o denotes the total number of data, the mathematical expression for the FCM objective function is equation (7):

$$\begin{gathered} J_{\min}(U,V) = \sum_{i_1=1}^{o} \sum_{j=1}^{c} u_{i_1 j}^{o^*} \left\| d_{i_1} - v_j \right\|_2^2 \\ u_{i_1 j} \geq 0, \sum_{j=1}^{c} u_{i_1 j} = 1, \forall i_1 \end{gathered} \tag{7}$$

where o^* denotes the fuzzy index. c denotes the total number of categories. i_1, j denote a certain sample and category. $u_{i_1 j}$ denotes the degree of affiliation of the j th sample to the i_1 th category. U denotes the affiliation matrix, V denotes the cluster center vector, and v_j is used to denote the cluster center vector where j is located.

Next, the Lagrange multiplier method is applied to realize the FCM objective function solution with the expression of Eq. (8):

$$J(U,V) = \sum_{i_1=1}^{o} \sum_{j=1}^{c} u_{i_1 j}^{o^*} \left\| d_{i_1} - v_j \right\|_2^2 + \sum_{i_1=1}^{c} \lambda_{i_1} \left(\sum_{j=1}^{c} u_{i_1 j} - 1 \right) \tag{8}$$

where λ_{i_1} denotes the Lagrange multiplier.

To find the minimum of the function $J_{\min}(U,V)$, the $u_{i_1 j}$, v_j partial derivatives in Eq. (7) are solved. Let $\frac{\partial J}{\partial U} = 0$, then we have Eq. (9):

$$o^* u_{i_1 j}^{o^*-1} \left\| d_{i_1} - v_j \right\|_2^2 - \lambda_{i_1} = 0 \tag{9}$$

Collation leads to equation (10):

$$u_{i_1 j} = \left(\frac{\lambda_{i_1}}{o^* \left\| d_{i_1} - v_j \right\|_2^2} \right)^{(o^*-1)^{-1}} \tag{10}$$

Due to $\sum_{j=1}^{c} u_{i_1 j} = 1$, we have equation (11):

$$\lambda_{i_1}^{(o^*-1)^{-1}} = \sum_{j}^{c}\left(o^*\left\|d_{i_1}-v_j\right\|_2^2\right)^{(o^*-1)^{-1}} \tag{11}$$

Combining Eq. (10) and Eq. (11) yields the affiliation update function, which can be expressed as Eq. (12):

$$u_{i_1 j} = \sum_{i_2=1}^{c}\lambda_{i_1 j}\cdot\left(\frac{\left\|d_{i_1}-v_j\right\|_2^2}{\left\|d_{i_1}-v_{i_2}\right\|_2^2}\right)^{(o^*-1)^{-1}} \tag{12}$$

Solving for its clustering center update function by making $\frac{\partial J}{\partial V}=0$, we have equation (13):

$$\sum_{i_1=1}^{o} u_{i_1 j}^{o^*}(d_{i_1}-v_j)=0 \tag{13}$$

Collation leads to equation (14):

$$v_j = \frac{\sum_{i_1=1}^{o}(u_{i_1 j})^{o^*} d_{i_1}}{\sum_{i_1=1}^{o}(u_{i_1 j})^{o^*}} \tag{14}$$

Through the above process, the U and V update formulas of the FCM algorithm can be derived. As a result, the process of multi-source heterogeneous data mining based on fuzzy C-mean clustering is as follows:

Input: multisource heterogeneous data interpolation processed dataset as equation (15):

$$D=\{d_1,d_2,\cdots,d_o\} \tag{15}$$

The number of clusters c, the fuzzy control index o^*, the maximum number of executions of the algorithm $t_{\max}$, and the termination threshold ξ.

The iterative operation process is as follows:

(1) Initialization: U constraints $u_{i_1 j}\geq 0$, $\sum_{j=1}^{c}u_{i_1 j}=1$, $\forall i_1$.

(2) Update the fuzzy partition matrix, i.e., the affiliation of the data, according to equation (12).

(3) Update the clustering center vector according to equation (14).

(4) Solve for the value of $J(U,V)$, if $J(U,V)\leq\xi$ then terminate the computation, otherwise, perform the next round of operations until $J(U,V)\leq\xi$.

Output: final clustering results to complete multi-source heterogeneous data mining.

IV. Performance Test and Evaluation of Multi-source Heterogeneous Data Mining Models

IV. A. Comparative validation of the effectiveness of data extraction methods

The governmental event distribution digital platform based on multiple data sources has a total of three data sources with an initial state of about 156,000 data, which are extracted using (K1) SSIS and (K2) this paper's model data extraction methods for the event distribution digital platform system based on multiple data sources. After the first extraction, the extraction is performed every 3 days and the time used for each extraction is recorded. Tables 1 and 2 show the time consumed for (K1) SSIS and (K2) this paper's model data extraction approach, respectively. It can be seen that the time consumed by the model data extraction approach in this paper is overall lower compared to the SSIS approach, and the time consumed in the 1st data extraction takes only 21.9mins.

Table 1: The consumption time of the K1 data extraction method

Times	Incremental data recording	Full data recording	Time-consuming
1	14.4W	14.3W	24.1mins
2	2.4W	15.8W	26.5mins
3	2.9W	17.8W	30mins
4	2.1W	19W	31.9mins
5	2.1W	20.2W	34mins
6	2.3W	21.6W	36.4mins
7	2.2W	22.9W	38.7mins
8	2.7W	24.7W	42.1mins

Table 2: The consumption time of the K2 data extraction method

Times	Incremental data recording	Full data recording	Time-consuming
1	14.4W	14.3W	21.9mins
2	2.4W	15.8W	23.5mins
3	2.9W	17.8W	26.3mins
4	2.1W	19W	27.9mins
5	2.1W	20.2W	29.8mins
6	2.3W	21.6W	31.9mins
7	2.2W	22.9W	33.8mins
8	2.7W	24.7W	36.6mins

A comparison of the time consuming data extraction for the two modeling approaches plotted in conjunction with the statistics in Tables 1 and 2 is shown in Figure 3.

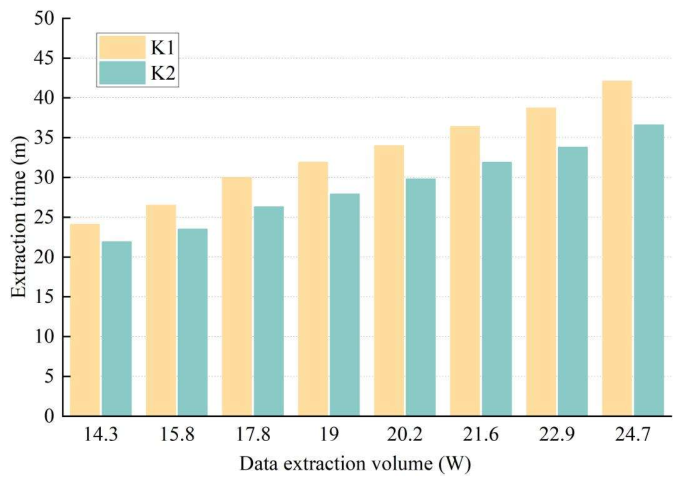

Figure 3: Comparison of data extraction time consumption of the two model methods

It can be seen that the time consumed by (K1) SSIS method data extraction is consistently more compared to (K2) model data extraction approach in this paper, with a difference of up to 5.5mins on the task of extracting data of 24.7 W. This indicates that (K2) model data extraction approach in this paper consumes less time for the same amount of data.

IV. B. Evaluation of the overall performance of the model

Comparing (S1) the multi-source heterogeneous data integration method proposed in this paper with (S2) the multi-source heterogeneous data integration system method utilizing big data technology with (S3) the key technology method of the urban geological survey all-space one map system, the system interface starts to read data from the storage space once it is successfully connected to the database. This process utilizes automatic extraction techniques in order to accurately refine the required data elements. All data execution and extraction is done in the background, ensuring that no detailed data is exposed during the extraction process. Once the extraction is complete, the results are transferred directly to the database, ensuring secure and efficient data processing. Experiments were conducted on three different data screening methods and the results are summarized in Figure 4.

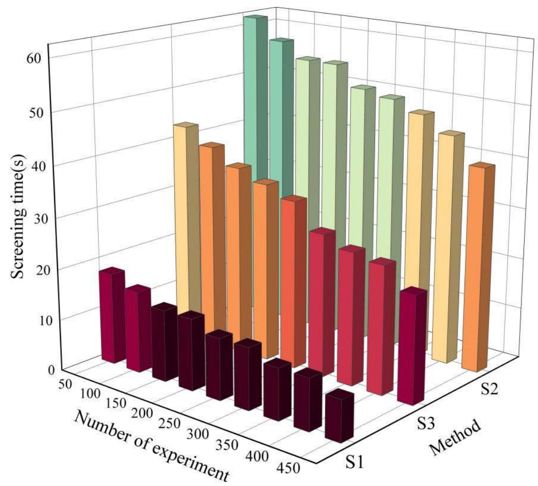

Figure 4: Comparison of data screening times by different methods

The data screening time of (S2) Multi-source Heterogeneous Data Integration System Method Using Big Data Technology and (S3) Key Technology Method of Urban Geological Survey Full Spatial One Map System are both longer, both of them are 20s and above. Compared with the two methods, (S1) the data screening time of this paper's method is shorter, not only lower than 20s but also as low as 8s in the 450th experiment, indicating that the multi-source heterogeneous data obtained by this paper's method is more accurate. In view of the diversified needs of government affairs in data management, different kinds of data can be extracted flexibly.

In order to verify the superiority of the multi-source heterogeneous data integration method proposed in this paper, the data integration performance test results of the method proposed in this paper are further compared with the two methods in Figure 5.

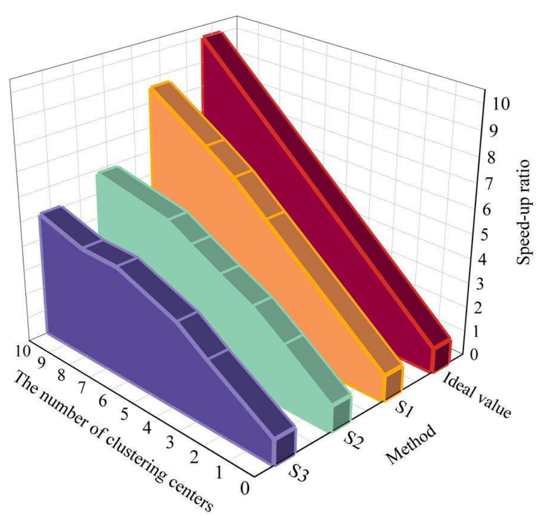

Figure 5: Comparison of speedup ratios under different methods

As seen in Fig. 5, (S1) the method proposed in this paper for multi-source heterogeneous data integration has a higher acceleration ratio and is closer to the ideal case compared to other methods. The highest acceleration ratio of (S1) this paper's method is around 8.59. In contrast, (S2) method and (S3) method have lower acceleration ratios, with the highest acceleration ratio of only 5.82.This indicates that this paper's method has advantages in multi-source heterogeneous data integration for government affairs.

In order to measure the applicability of this paper's method more deeply, data mining integration comparative experiments are conducted using three different methods. The stability test results are shown in Table 3.From Table 3, it can be seen that at 30 min, (S1) this paper's method achieves 97.93% stability, while the stability of (S2) method and (S3) method is 83.73% and 73.73% respectively. And with the increase of time, the stability of (S1) method of this paper maintains a high level, always in the range of 96.00%-8.00%, and all of the other two methods fluctuate more. By analyzing the data results in detail, it can be clearly seen that the method of this paper has a clear advantage in terms of stability and performs better compared with the other two methods.

Table 3: Stability comparison of different methods

Time(min)	S1(%)	S2(%)	S3(%)
30	97.93	83.73	73.73
60	97.43	83.63	73.43
90	96.73	83.43	74.03
120	97.63	83.43	73.63
150	97.03	84.53	77.03
Average	97.35	83.75	74.37

V. Conclusion

In this paper, we use the key technology framework of multi-source heterogeneous data fusion for sensitive data to process and capture multi-source heterogeneous data in government events. A classifier combining log text classifier and context similarity comparator is designed for change log data capture, and the proposed multi-source heterogeneous data mining model is synthesized. The proposed model consumes a minimum of 21.9mins for data extraction and has a maximum difference of 5.5mins with similar modeling methods, and the data screening time is consistently under 20s under various experimental conditions with an acceleration ratio of up to 5.82. The stability of the data integration not only reaches a maximum of 97.93%, but also stays stable between 96.00% and 98.00%. 98.00%.

By using the government event distribution platform based on the event distribution model to collect and distribute multiple government events, combined with the multi-source heterogeneous data mining model, the processing speed of government events can be effectively improved, providing a reliable reference for the optimization and improvement of administrative efficiency under the digital government platform.

Funding

Study on the Conceptual Transformation of National Ocean Governance and Modernization of Ocean Governance in the South China Sea.

References

[1] Steinebach, Y. (2023). Administrative traditions and the effectiveness of regulation. Journal of European Public Policy, 30(6), 1163-1182.
[2] Khan, A. H. (2017). Administrative efficiency and effectiveness with the application of e-government: A study on Bangladesh public administration. In User centric e-government: Challenges and opportunities (pp. 105-116). Cham: Springer International Publishing.
[3] Dike, E. E. (2019). E-governance and administrative efficiency: Issues and challenges. International Journal of Innovative Research in Education, Technology & Social Strategies, 6(1), 184-194.
[4] Doran, N. M., Puiu, S., Bădîrcea, R. M., Pirtea, M. G., Doran, M. D., Ciobanu, G., & Mihit, L. D. (2023). E-government development—A key factor in government administration effectiveness in the European Union. Electronics, 12(3), 641.
[5] Zhang, L., & Zhang, X. (2025). Impact of digital government construction on the intelligent transformation of enterprises: Evidence from China. Technological Forecasting and Social Change, 210, 123787.
[6] Dong, C., Liu, J., & Mi, J. (2023). How to enhance data sharing in digital government construction: A tripartite stochastic evolutionary game approach. Systems, 11(4), 212.
[7] Liu, F., Liu, G., Wang, X., & Feng, Y. (2024). Whether the construction of digital government alleviate resource curse? Empirical evidence from Chinese cities. Resources Policy, 90, 104811.
[8] Wang, S., Sun, X., & Zhong, S. (2023). Exploring the multiple paths to improve the construction level of digital government: Qualitative comparative analysis based on the WSR framework. Sustainability, 15(13), 9891.
[9] Chen, B. (2021, March). Research on the construction of digital government in Digital Economy. In 6th International Conference on Financial Innovation and Economic Development (ICFIED 2021) (pp. 351-356). Atlantis Press.
[10] Tao, S., Liu, H., Wang, S., & Li, C. (2020). Construction of smart coastal cities based on digital government. Journal of Coastal Research, 110(SI), 154-158.
[11] Xu, H. (2024). How does digital government affect energy efficiency?. Management of Environmental Quality: An International Journal, 35(7), 1524-1544.
[12] Yang, C., Gu, M., & Albitar, K. (2024). Government in the digital age: Exploring the impact of digital transformation on governmental efficiency. Technological Forecasting and Social Change, 208, 123722.
[13] Gao, B., & Jin, Q. (2024). The Pathway for Modernization Transformation in Government Governance——Construction of Digital Intelligent Government. Economics & Management Review, 5(1).
[14] Wibawa, R. P. (2024). Digital transformation and administrative efficiency study. American Journal of Science and Learning for Development, 3(1), 169-175.