Exploring the Evolution of Modern Chinese History through Digital Method and Data Mining

Huaili Hu[1*]

[1]School of Teacher, Anhui University of Science & Technology, Huainan, Anhui, 232001, China

Corresponding Author: Huaili Hu, ah8655285111@163.com

Abstract. In order to promote the effect of historical evolution data analysis, this paper combines data mining algorithms to analyze the evolution of modern Chinese history, and improves the effect of modern history evolution analysis through intelligent models. Furthermore, this paper proposes a blind recognition algorithm based on scrambling-cancelled composite codes. For the identification and analysis of complex codes with different inter-frame scrambling codes, a synchronous blind identification algorithm for complex codes is proposed in this paper. In addition, according to the actual situation of historical evolution data analysis, this paper proposes a method for analyzing the evolution of modern Chinese history based on data mining. The simulation results show that the performance of the improved algorithm has been significantly improved compared with the original algorithm, and the experimental results verify that the evolution method of modern Chinese history based on data mining proposed in this paper can play an important role in historical data processing.

Keywords: data mining; China; modern history; evolution analysis;
DOI: https://doi.org/10.14733/cadaps.2024.S16.286-308

1 INTRODUCTION

Knowledge storage is to store the acquired triple data in a unified format. At present, there are two main storage methods of knowledge graph: RDF file method and graph database method. Using RDF files to store data, the constructed data model has graph characteristics. However, when faced with a huge amount of RDF file data in practical applications, using traditional data management methods, namely relational data model or object data model, it is difficult to meet the two requirements of low data redundancy and high query performance at the same time. Since reducing data redundancy will increase relational connections and reduce query performance, more relational connections are required to improve query performance, resulting in data redundancy [5].

When RDF data is managed graphically, the conversion between RDF logical data model and physical data model can be effectively avoided, and the query performance can be improved by using graph theory related algorithms. When graph database is used to store knowledge atlas, entities and relationships between entities become subjects. Graph structured data model stores entities with vertices of graph and relationships between entities with edges. Compared with traditional relational databases, this data model has higher computing efficiency [15] and storage efficiency for large, complex, interconnected, and variable types of data, and can also maintain good efficiency when traversing large amounts of data. The commonly used graph databases are Neo4j, OrientDB, GraphDB, etc. The Neo4j graph database is the most commonly used one in scientific research or engineering. Neo4j database uses its own unique cypher [11] language for data query. At the same time, Neo4j database has a good visualization effect. There are a large number of entities and complex relational data in the knowledge map. Therefore, compared with the relational database, Neo4j map database does not need to design the data structure before processing data because its data model is descriptive, which greatly improves the flexibility and convenience of the knowledge map in the data expansion process. Taking interpersonal relationship as an example, when describing this relationship, the relational database needs to design two tables to store the attributes of people and interpersonal relationships respectively. When adding people and relationships, you need to operate two tables. If there is any addition or deletion of attributes, you also need to change the structure of the table, which is inefficient. In the graph database, there is no table structure limitation. When expanding data, you only need to add corresponding nodes or edges, and the related attribute descriptions only need to operate on a single node or edge [10]. Gather digital data related to modern Chinese history. This may include historical documents, photographs, artworks, audio recordings, and more. Access to digital archives, libraries, museums, and government databases can be valuable sources

As the carrier of social culture, historical archives highlight the unique art, custom, morality and belief of the Chinese nation. The reasonable development of historical archives can not only meet the needs of the public for spiritual civilization, but also have unique advantages in promoting excellent traditional culture. First of all, rich historical records enable contemporary people to directly and truly understand history. All historical archives collection institutions give full play to their functions of cultural heritage and leisure through the development and utilization of archives. With the help of various forms such as compilation and research of archives and historical materials and archives exhibition, the public can obtain historical knowledge through a large number of text, pictures and other archives resources, and feel the broad and profound Chinese culture. Secondly, respect for historical archives is the premise and foundation of cultural creation, literary creation and literary creation. Historical archives always reflect the truth, goodness and beauty of literary creation [17]

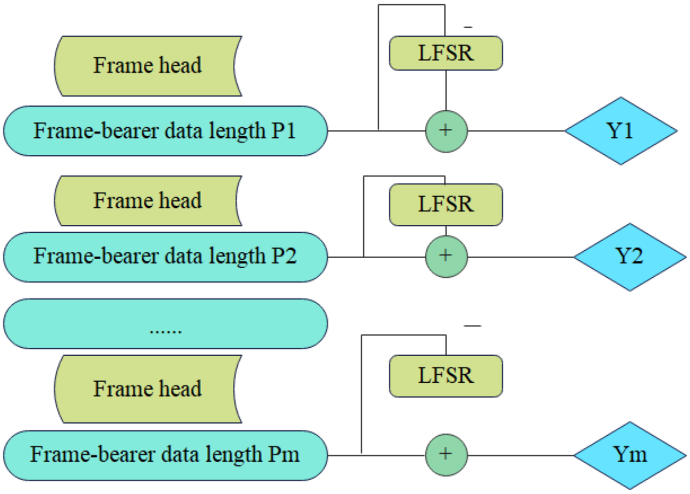
Under the background of the digital era, the development and utilization of historical archives are faced with the new demand of digitalization, knowledge and intelligence. Digital requirements. From the perspective of resource construction, the types of historical archives resources are diverse, including text, image, audio and video and other carrier forms. Digital processing and transformation of the above archives by means of digital means has obvious epochal significance for strengthening the digital storage, management, retrieval and presentation of historical archives. From the perspective of archive users, the growing digitization demand of archive users is not only the external representation of their own archive utilization rules, but also an important target for quality control in the process of archive digitization planning [12]. Knowledge and intelligence requirements. The characteristics of archive users' demand for historical archives derive new changes on the basis of digitalization, and tend to be more knowledge-based and intelligent. Users' demands for knowledge and intelligence depend on the knowledge processing of archive resources. The digitalization of historical archives resources is an important basis and premise for library collection knowledge processing. Through the knowledge organization and knowledge processing of archives data resources, the goal of knowledge aggregation is further achieved, and the knowledge products with more fine granularity can meet the knowledge and intelligence needs of users. At present, traditional one-dimensional retrieval is still mostly used for the aggregation of historical archives resources. This mode is characterized by its simplicity and convenience. It does not need to know about the system or learn relevant retrieval technologies, but it is difficult to meet users' multi-level needs for information utilization [1]. The traditional resource management mode is to store different types of historical archives resources in specific thematic databases according to the theme. Although this approach can achieve the integration of historical archives resources of different media types, it also causes the internal links between historical archives resources to be separated, which is not conducive to further in-depth resource development. Historical archives resources that have not been aggregated can only be submerged in the information ocean [14]. At this stage, archives should meet users' needs through various channels and channels. The data barrier of archive resources needs more practical and operable technologies to break through, break through the shortcomings of traditional knowledge organization, overcome the complexity and heterogeneity of archive data, and achieve semantic interconnection [8]. In general, associated data has significant advantages in the aggregation of historical archives resources, mainly in four aspects: establishing semantic links with the largest association strength, breaking domain barriers, aggregating historical archives resources with different association strengths, and making semantic expression clearer [7]. The associated data technology system is used to solve the problems of semantic organization, semantic interconnection, visual analysis, etc. of historical archives resources, and provide solutions to meet the new needs of archive users for the utilization of historical archives resources that are changing constantly [9].

The application of modern information technology in the field of archives management has been developing in depth, and new technologies such as big data, cloud computing, blockchain, artificial intelligence, digital humanities have also been applied in the development process of historical archives resources [3]. Under the background of rapid development of new technologies, historical archives resources rely on the development and utilization of network technology, computer technology and multimedia technology, and have undergone tremendous changes in both ways and scales. Although the development and utilization of historical archives has received unprecedented attention and promotion, there are still some problems that have not been solved in depth [13]. "Information island" phenomenon. Under the background of rapid digital growth of archive resources, the collection units, institutions and departments of various historical archives at all levels lack communication and contact, and are relatively independent in management. In the process of archive resource management, most of the archive information networks are still limited, and the historical archive resources of different collection institutions are not related to each other, which is in conflict with users' demand for comprehensive access to historical resources, Historical archives resources are in the state of "information island", and there are certain difficulties in the development, utilization and sharing of many historical archives resources, which greatly limits the application value of historical archives [4]. "Heterogeneous resources". In the process of archive data, historical archive data has different format standards and different knowledge organization schemes. Archival data includes metadata records, knowledge organization systems, historical background information and many other resources. Archival data is not only structured data stored in documents or relational databases, but also semi-structured and unstructured data stored in Word, PDF, HTML and other formats. Therefore, the original data format should be converted, and the semantic conversion of documents, forms, XML, terminology and knowledge organization systems should be conducted for different data types. Although metadata can provide a semantic basis for resources, it cannot completely solve the problem of semantic heterogeneity of information systems [16]. Using different metadata schemes at the same time will result in heterogeneous resources or complex associations between resources [2]. Although thesaurus and taxonomy have hierarchical structure and classification level, their expressions are too simple and conservative for the moment. Transforming archive data resources into associated data format through semantics can realize the sharing and reuse of archive domain knowledge to a certain extent [6].

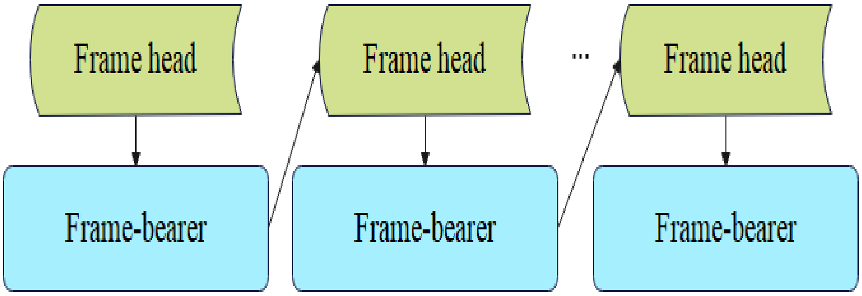
This paper combines the data mining algorithm to analyze the evolution of modern Chinese history, and improves the analysis effect of the evolution of modern history through the intelligent model.

2 COMPLEX CODE IDENTIFICATION OF CHANNEL CODING AND SCRAMBLING CODE FOR HISTORICAL INFORMATION

The commonly used synchronous scrambling structure in communication system is shown in Figure 1(a). For the fixed frame length, the length of each frame is fixed, and the length and content of the frame header of each frame are also fixed. For this structure, the transmission data frame structure can be represented as Figure 1(b).



(a) Scrambling method of communication system



(b) Fixed frame header data transmission structure

Figure 1: Communication system structure.

Simulation conditions are set: the frame header is "000001100101", the frame length is 256, 50 frames, and the data is cyclically shifted to the left by 7 bits to simulate the actual situation, and the bit error rate is 0.001. The simulation results are shown in Figure 2.

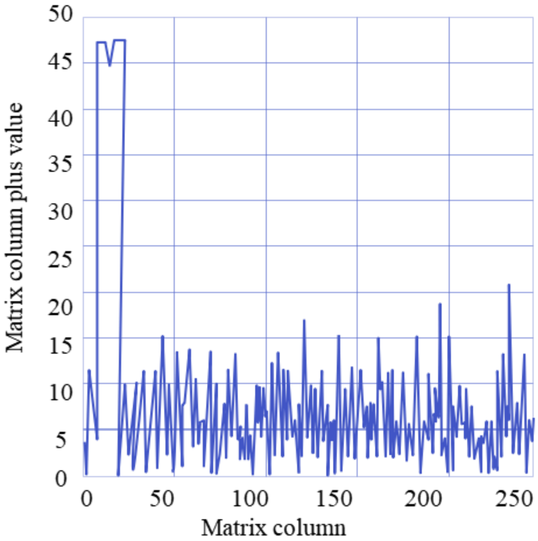


Figure 2: Direct estimation method of fixed frame length.

It can be seen from the simulation results that the detection method is effective. Due to the existence of bit errors, the peak value of the frame header will fluctuate slightly, but it has obvious advantages compared with other accumulators. The corresponding starting point of the peak is the starting position of the first complete received frame in the data stream, and the length of the continuous peak is the length of the frame header.

Simulation condition setting: TM frame header of CCSDS, the decimal representation is "56081C971AA73D3E", 13 frames, and the bit error rate is 0.0003. The start of each identified frame is "13216611021140118012221260131213601410146215161", and then divided into frames based on the position of the identified frame start in the data stream, and each frame is truncated and transformed into a matrix with the minimum frame length, as shown in Figure 3.

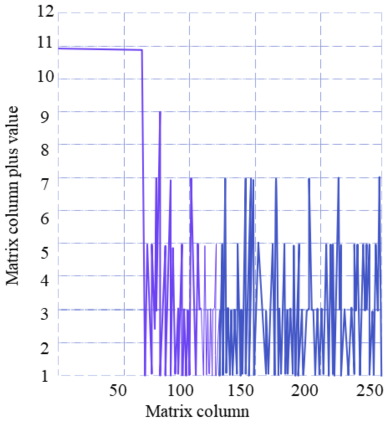


Figure 3: Correlation search method.

It can be seen from the simulation results that the method is effective for frame synchronization recognition with variable frame length.

2.1 Complex Code Identification of Scrambling Code and Channel Coding of Historical Information

According to the structure of the scrambler in Fig. 1, the scrambling code is canceled by using the truncation property that the scrambling position of the frame is the same and the scrambling sequence is the same scrambling code sequence. If it is assumed that there are two frames of received frame body data R_1, R_2 after removing the frame header, and the length relationship between the two frames is $R_1 \le R_2$,the corresponding information sequence and scrambling code sequence are M_1, Cr_1 and M_2, Cr_2 respectively, and their relationship can be expressed as:

$$\begin{aligned} R_1 &= M_1 \oplus Cr_1 \\ R_2 &= M_2 \oplus Cr_2 \end{aligned} \tag{1}$$

The scrambling sequences corresponding to the two frames are generated by the same LFSR structure and register initial state, but their lengths are different. Truncating R_2 according to the length of R_1, the corresponding information sequence is M_2'. At this time, the two sequences of R_1 and R_2' can be expressed as:

$$\begin{aligned} R_1 &= M_1 \oplus Cr_1 \\ R_2' &= M_2' \oplus Cr_1 \end{aligned} \tag{2}$$

The XOR operation of the two output sequences can be obtained:

$$\begin{aligned} R &= R_1 \oplus R_2' \\ &= \left(M_1 \oplus Cr_1\right) \oplus \left(M_2' \oplus Cr_1\right) \\ &= M_1 \oplus M_2' \end{aligned} \tag{3}$$

The specific algorithm flow is as follows:

1. The algorithm sorts the frames according to the frame length from small to large, and counts the number of frames of each frame length, and assumes that the statistical result is $n_1, n_2, n_3, \cdots, n_k$.

2. According to the frame length from small to large, the algorithm extracts the frame bearers of two frames from the frame length whose statistical frame length is greater than 1 each time and puts them into two sequences respectively, and S_1、S_2, f_n、f_n represents different frames of the same frame length, respectively, where $s \le k$:

$$S_1 = \left(f_{n_1}, f_{n_2}, \cdots, f_{n_s} \right)$$
$$S_2 = \left(f_{n_1}', f_{n_2}', \cdots, f_{n_s}' \right) \quad (4)$$

3. The algorithm performs XOR operation on the constructed two sequences S_1 and S_2 to obtain the sequence $y, y = S_1 \oplus S_2$.

1. Scrambling cancellation and identification simulation of linear block codes

The simulation conditions are set: (7,4) Hamming code, the generator matrix is: $G = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$, the starting point of the code word is 1; the generator polynomial of the scrambling code is $g(x) = x^8 + x^4 + x^3 + x^2 + 1$, the initial state is "11111111", and the amount of data is 3500. The identification result is shown in Figure 4. The bit error rate in the figure is XORed to get the bit error rate setting of the codeword.

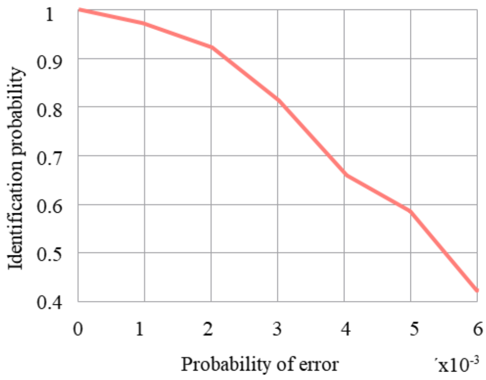


Figure 4: The relationship between the recognition probability of hamming code of scrambling code cancellation and the bit error rate.

2. Scrambling code cancellation and identification simulation of convolutional code

The convolutional coding adopts the 1/2 code rate, the constraint degree is 7, and the generating matrix is the convolutional code of $[133 \quad 171]$. The generator polynomial of the scrambling code is $g(x) = x^8 + x^4 + x^3 + x^2 + 1$, the initial state is "11111111", the amount of data is 3500, and the number of experiments is 100 Monte Carlo experiments. The two data sequences are aligned after the convolution encoding of the above parameters, and the XOR operation is performed to perform convolution encoding identification on the obtained data. The relationship between the recognition probability and the bit error rate is shown in Figure 5. The bit error rate in the figure is XORed to obtain the bit error rate setting of the codeword, and the bit error rates of the two sequences are equal. The sequence obtained by the combination of the convolutional code and the scrambling code is still recognized as the convolutional code after the scrambling code is cancelled.

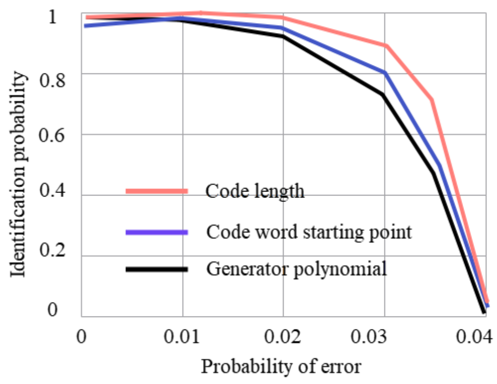


Figure 5: Relationship between recognition probability of convolutional coding with scrambling cancellation and bit error rate.

3. Scrambling code cancellation and identification simulation of deleting convolutional code

The simulation conditions are set: 2/3 code rate, the constraint degree is 7, the generator matrix is [133171], and the deletion mode is [001110] deletion convolutional code. Since the amount of data required to delete the convolutional code is more than that of the convolutional code, the amount of data is 7000, and the number of experiments is 100 Monte Carlo experiments. The recognition result is shown in Figure 6:

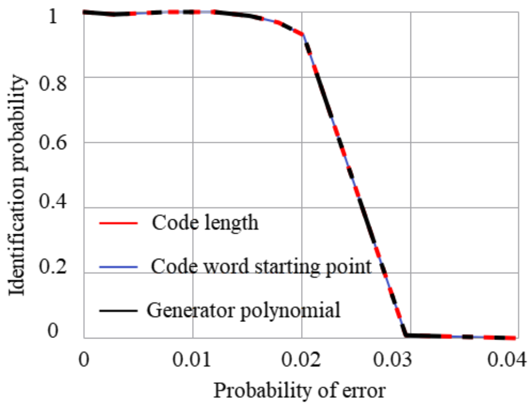


Figure 6: Relationship between the recognition probability of deleting convolutional code and the bit error rate.

The scrambling code parameter identification algorithm based on the dual codeword utilizes the orthogonality between the dual codeword and the coded codeword and the shifting and sampling characteristics of the m-sequence to complete the identification of the scrambling code.

Simulation of scrambling code recognition of dual codewords of linear block codes

Simulation conditions are set: (7,4) Hamming code, the LFSR generator polynomial is: $P(X)=X^8+X^4+X^3+X^2+1$, the initial state is "11111111", and the check matrix is:

$$H=\begin{bmatrix}1&0&1&1&1&0&0\\1&1&1&0&0&1&0\\0&1&1&1&0&0&1\end{bmatrix} \tag{3}$$

The identification results are as follows: $Q(X): X^{112}+X^7+1, X^{224}+X^{14}+1, X^{266}+X^{245}+1$. The second one is the multiple of the first polynomial, so the multiple of the generator polynomial is $X^{112}+X^7+1, X^{266}+X^{245}+1$. The greatest common factor of a multiple polynomial is $X^{56}+X^{42}+X^{35}+X^{21}+1$. By factoring the greatest common factor, we get (1) $X^{24}+X^{20}+X^{16}+X^{15}+X^{14}+X^{13}+X^{12}+X^{11}+X^7+X^6+X^5+X^4+X^3+1$,(2) $X^{24}+X^{19}+X^{18}+X^{16}+X^{14}+X^{13}+X^{11}+X^6+X^4+X^2+1$, (3) $X^8+X^4+X^3+X^2+1$.

We assume that the identified linear block code parameters are (n, k), and the parity check matrix is H. The received data is $y=(y_1, y_2, \cdots)$, and a matrix is constructed for the received data according to the code length n, where N is the received sequence length, $y_m=\left(y_{(m-1)^*n+1}, y_{(m-1)^*n+2}, \cdots, y_{m^*n}\right)$:

$$Y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix}=\begin{bmatrix}y_1&y_2&\cdots&y_n\\y_{n+1}&y_{n+2}&\cdots&y_{2^*n}\\\vdots&\vdots&\vdots&\vdots\\y_{m^*n+1}&y_{m^*n+2}&\cdots&y_{(m+1)^*n}\end{bmatrix} \tag{6}$$

When the combined decimation is set for the Y matrix, the corresponding polynomial is $Q(X)=1+C_{i_1}X^{n_i}+C_{i_2}X^{n_2}+\cdots+C_{i_{d-1}}X^{n_{d-1}}$:

$$\begin{aligned}C_t' &= y_t \oplus y_{t-i_1} \oplus \cdots \oplus y_{t-i_{d-1}}\\ &=\left(C_t \oplus m_t\right) \oplus \left(C_{t-i_1} \oplus m_{t-i_i}\right) \oplus \cdots \oplus \left(C_{t-i_{d-1}} \oplus m_{t-i_{d-1}}\right)\\ &=\left(C_t \oplus C_{t-i_1} \oplus \cdots \oplus C_{t-i_{d-1}}\right) \oplus \left(m_t \oplus m_{t-i_1} \oplus \cdots \oplus m_{t-i_{d-1}}\right)\end{aligned} \tag{7}$$

$C_{t-i}=\left(c_{t-i,0}, c_{t-i,1}, \cdots, c_{t-i,n-1}\right)$ represents a complete codeword corresponding to the t-ith row of the matrix Y, and $m_{t-i}=\left(m_{t-i,0}, m_{t-i,1}, \cdots, m_{t-i,n-1}\right)$ represents the scrambling code sequence corresponding to the t-ith row of the matrix Y. The statistical components are:

$$z_t = max\left(C_t{}'H^T\right) \tag{8}$$

Then the statistic is:

$$Z = \sum_{j=i_{d-1}}^{N} (-1)^{it} \tag{9}$$

The simulation conditions are set: (7,4) Hamming code, the LFSR generator polynomial is $P(X) = X^7 + X + 1$, the initial state is "1111111", and the check matrix is $H = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$.

The identification results are as follows: $Q(X): X^{133} + X^{126} + 1, X^{175} + X^{119} + 1, X^{182} + X^{140} + 1\$,\$X^{266} + X^{252} + 1$. The fourth is the multiplication of the first polynomial, so the multiplication of the generator polynomial is $X^{133} + X^{126} + 1, X^{175} + X^{119} + 1, X^{182} + X^{140} + 1$. The greatest common factor of a multiple polynomial is $X^{49} + X^{42} + X^{35} + X^{28} + X^{21} + X^{14} + 1$. By factoring the greatest common factor, we get (1) $X^{21} + X^{20} + X^{19} + X^{15} + X^{11} + X^9 + X^7 + X^3 + X + 1$, and the initial state is "0000000";(2) $X^{22} + X^{14} + X^9 + X^7 + X^3 + X^2 + 1$, and the initial state is "0000000"; (3) $X^7 + X + 1$, and the initial state is "1111111". It can be known from the initial state of the LFSR register that the sequence of all 0s cannot be used, and the third factor meets the requirements and completes the identification.

2.2 Synchronous Blind Recognition of Historical Information Channel Coding and Scrambling Codes with Different Interframe Scrambling Codes

As shown in Figure 7, when the corresponding polynomial Q(X) is extracted in combination, and the LFSR generator polynomial P(X) satisfies Q(X)=A(X)P(X), the historical information channel coding is performed on the constructed s sequence. When the information channel coding parameters are used, the historical information channel coding parameters and Q(X) are determined at the same time.

$$\begin{aligned} s_t &= y_t \oplus y_{t-i_1} \oplus \cdots \oplus y_{t-i_{d-1}} \\ &= \left(C_t \oplus m_t\right) \oplus \left(C_{t-i_1} \oplus m_{t-i_i}\right) \oplus \cdots \oplus \left(C_{t-i_{d-1}} \oplus m_{t-i_{d-1}}\right) \\ &= \left(C_t \oplus C_{t-i_1} \oplus \cdots \oplus C_{t-i_{d-1}}\right) \oplus \left(m_t \oplus m_{t-i_i} \oplus \cdots \oplus m_{t-i_{d-1}}\right) \end{aligned} \tag{10}$$

When Q(X)=A(X)P(X), that is, the m sequence generated by the generator polynomial of Q(X) is the cyclic shift sequence of the m sequence generated by P(X), so there are:

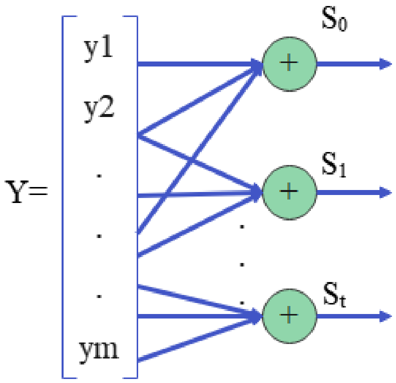


Figure 7: The structure of the sequence s constructed from the matrix Y.

$$\left(\mathrm{m}_t \oplus \mathrm{m}_{t-i_1} \oplus\cdots\oplus \mathrm{m}_{t-i_{d-1}}\right)^T = \begin{bmatrix} m_{t,0}\oplus m_{t-i_1,0}\oplus\cdots\oplus m_{t-i_{d-1},0} \\ m_{t,1}\oplus m_{t-i_1,1}\oplus\cdots\oplus m_{t-i_{-1},1} \\ \vdots \\ m_{t,n-1}\oplus m_{t-i_1,n-1}\oplus\cdots\oplus m_{t-i_{d-1},n-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{11}$$

Then,

$$\mathrm{s}_t = \mathrm{y}_t \oplus \mathrm{y}_{t-i_1} \oplus\cdots\oplus \mathrm{y}_{t-i_{d-1}} = \mathrm{C}_t \oplus \mathrm{C}_{t-i_1} \oplus\cdots\oplus \mathrm{C}_{t-i_{d-1}} \tag{12}$$

$$\mathrm{s} = \left(\mathrm{s}_1, \mathrm{s}_2, \cdots, \mathrm{s}_t, \cdots\right) \tag{13}$$

The complex code of historical information channel coding and scrambling code Historical information channel coding parameters and generating polynomial multiplication parameters identification algorithm flow is as follows:

Step1: The algorithm sets the code length traversal range $n_1 \le n \le n_2$.

Step2: The algorithm divides the received sequence $\mathrm{y} = \left(y_1, y_2, \cdots\right)$ into blocks, each block contains n-bit data of code length, and aligns each block into a matrix Y;

$$\mathrm{Y} = \begin{bmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \\ \vdots \\ \mathrm{y}_m \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \\ y_{n+1} & y_{n+2} & \cdots & y_{2^* n} \\ \vdots & \vdots & \vdots & \vdots \\ y_{m^* n+1} & y_{m^* n+2} & \cdots & y_{(m+1)^* n} \end{bmatrix} \tag{14}$$

Step3: The algorithm sets different combinations of extraction.

Step 4: The algorithm extracts the current combination of the matrix Y to construct the sequence s, and the corresponding polynomial of the combination is Q(X), and the sequence is identified by channel coding of historical information. If the channel coding parameters of the historical information are identified, it is judged whether Q(X) is the same as the stored polynomial. If they are different, the algorithm stores Q(X), if they are the same, it does not store it. The algorithm judges that the combination traversal is over. If yes, the algorithm goes to Step5, otherwise it goes to Step3.

Step5: The algorithm obtains the greatest common divisor B(X) of the stored multi-forms by the method of tossing and turning.

Step6: The algorithm uses Berlekamp algorithm to factorize the greatest common factor B(X), $B(X) = B_1(X) \cdot B_2(X) \cdots$.

We assume that the generator polynomial of LFSR is $P(X) = 1 + a_1 x + a_2 x^2 + \cdots + a_L x^L, a_i \in \{0,1\}, L$ is the series of generator polynomials, the identified historical information channel coding parameters are (n, k), the codeword is C, and the check matrix is H:

$$\mathrm{H} = \begin{bmatrix} h_{0,0} & h_{0,1} & \cdots & h_{0,n-1} \\ h_{1,0} & h_{1,1} & \cdots & h_{1,n-1} \\ \vdots & \vdots & \vdots & \vdots \\ h_{n-k-1,0} & h_{n-k-1,2} & \cdots & h_{n-k-1,n-1} \end{bmatrix} \tag{15}$$

$\mathrm{h}_0, \mathrm{h}_1, \cdots \mathrm{h}_{n-k-1}$ is the H-th row of $0,1,\cdots,n-k-1$, which are the dual codewords of codeword C. The received sequence is $\mathrm{y}_t = (y_0, y_1, \ldots)$, and $\mathrm{Y}_0, \mathrm{Y}_1, \ldots$ is obtained by dividing it into blocks according to the channel coding code length n of the historical information, then the i-th block can be expressed as:

$$\mathrm{Y}_i = \left(y_{ni}, y_{ni+1}, \cdots y_{(n+1)i-1}\right) \tag{16}$$

The constructed $\mathrm{r} = (r_0, r_1, r_2, \cdots)$-sequence is shown in Figure 8:

$$\begin{aligned} r_0 &= \mathrm{Y}_0 \mathrm{h}_0^T = \sum_{i=0}^{n-1} y_i \cdot h_{0,i} \\ &= y_0 \cdot h_{0,0} \oplus y_1 \cdot h_{0,1} \oplus \cdots \oplus y_{n-1} \cdot h_{0,n-1} \\ r_1 &= \mathrm{Y}_1 \mathrm{h}_0^T = \sum_{i=0}^{n-1} y_{n+i} \cdot h_{0,i} \\ &= y_n \cdot h_{0,0} \oplus y_{n+1} \cdot h_{0,1} \oplus \cdots \oplus y_{2n-1} \cdot h_{0,n-1} \\ &\vdots \end{aligned} \tag{17}$$

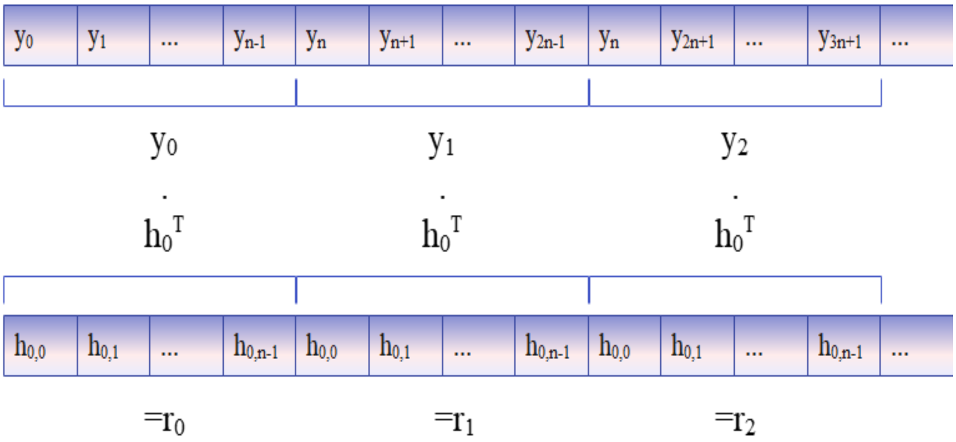


Figure 8: The Structure of the Linear Block Code Construction Sequence r.

Because of $y_t = c_t \oplus s_t, \mathrm{Ch}_0^T = 0$, the constructed r can be written as:

$$\begin{aligned} r_i &= \mathrm{Y}_i \mathrm{h}_0^T = \mathrm{C}_i \mathrm{h}_0^T \oplus \mathrm{S}_i \mathrm{h}_0^T \\ &= \mathrm{S}_i \mathrm{h}_0^T \end{aligned} \tag{18}$$

$$\begin{aligned} r_0 &= s_0 \cdot h_{0,0} \oplus s_1 \cdot h_{0,1} \oplus \cdots \oplus s_{n-1} \cdot h_{0,n-1} \\ r_1 &= s_n \cdot h_{0,0} \oplus s_{n+1} \cdot h_{0,1} \oplus \cdots \oplus s_{2n-1} \cdot h_{0,n-1} \\ &\vdots \end{aligned} \tag{19}$$

Among them, C_i represents the codeword corresponding to the i-th block, and S_i represents the m sequence corresponding to the i-th block. The output of LFSR at time t is:

$$s_t = \sum_{i=1}^{L} a_i s_{t-i} \tag{20}$$

If the state of the LFSR at time t is assumed to be $\mathrm{S}_t = \left(s_t s_{t+1} s_{t+2} \cdots s_{t+L-1}\right)^T$, the transition matrix F is defined as:

$$\mathrm{F} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & a_{L-1} & a_{L-2} & \cdots & a_2 & a_1 \end{pmatrix} \tag{21}$$

The state of the LFSR register at the $t+i, i=0,1,2,\cdots$ -th moment is: $S_{t+i} = F^i S_t$. If $G \in L \times L, U \in 1 \times L$ and $U = (1 \quad 0 \quad 0 \quad \cdots \quad 0)$, then there is $s_t = US_t = UF^t S_0$. From formula (17), we get:

$$\begin{aligned} r_0 &= US_0 \cdot h_{0,0} \oplus US_1 \cdot h_{0,1} \oplus \cdots \oplus US_{n-1} \cdot h_{0,n-1} \\ &= U\left(I_L \cdot h_{0,0} \oplus F \cdot h_{0,1} \oplus \cdots \oplus F^{n-1} \cdot h_{0,n-1}\right) S_0 \\ r_1 &= U\left(I_L \cdot h_{0,0} \oplus F \cdot h_{0,1} \oplus \cdots \oplus F^{n-1} \cdot h_{0,n-1}\right) F^n S_0 \\ r_2 &= U\left(I_L \cdot h_{0,0} \oplus F \cdot h_{0,1} \oplus \cdots \oplus F^{n-1} \cdot h_{0,n-1}\right) F^{2n} S_0 \\ &\vdots \\ r_i &= U\left(I_L \cdot h_{0,0} \oplus F \cdot h_{0,1} \oplus \cdots \oplus F^{n-1} \cdot h_{0,n-1}\right) F^{in} S_0 \end{aligned} \tag{22}$$

We define a G matrix and denote it as:

$$G = \begin{pmatrix} U\left(I_L \cdot h_{0,0} \oplus F \cdot h_{0,1} \oplus \cdots \oplus F^{n-1} \cdot h_{0,n-1}\right) I_L \\ U\left(I_L \cdot h_{0,0} \oplus F \cdot h_{0,1} \oplus \cdots \oplus F^{n-1} \cdot h_{0,n-1}\right) F^n \\ \vdots \\ U\left(I_L \cdot h_{0,0} \oplus F \cdot h_{0,1} \oplus \cdots \oplus F^{n-1} \cdot h_{0,n-1}\right) F^{(L-1)n} \end{pmatrix} \tag{23}$$

Therefore, the initial state is obtained as:

$$S_0 = G^{-1} \left(r_0 \quad r_1 \quad \cdots r_{L-1} \right)^T \tag{24}$$

2.3 Analysis on the Blind Recognition Algorithm of Composite Codes Based on Channel Coding and Scrambling of Historical Information Under Bit Errors

n represents the coding code length of the historical information channel, the data received at time t is y_t^e, the corresponding error pattern is $e_t = \left(e_{nt}, e_{nt+1}, \cdots, e_{(n+1)t-1}\right)$, and the error pattern refers to the error code introduced in the transmission.

The output can be expressed as:

$$y_t^e = y_t \oplus e_t = c_t \oplus s_t \oplus e_t = s_t \oplus \left(c_t \oplus e_t\right) \tag{25}$$

$$y_t^e = y_t \oplus e_t \tag{26}$$

Because of $y_t = c_t \oplus s_t$, the received data is multiplied by the dual codeword in blocks of code length n:

$$r_i^e = Y_i^e \mathrm{h}_0^T = C_i \mathrm{h}_0^T \oplus S_i \mathrm{h}_0^T \oplus E_i \mathrm{h}_0^T \tag{27}$$

Among them, E_i represents the error pattern corresponding to the ith block, and because of $\mathrm{C}_i \mathrm{h}_0^T = 0$,

$$r_i^e = \mathrm{S}_i \mathrm{h}_0^T \oplus \mathrm{E}_i \mathrm{h}_0^T \tag{28}$$

$$\begin{aligned}
r_0^e &= (s_0 \oplus e_0) \cdot h_{0,0} \oplus (s_1 \oplus e_1) \cdot h_{0,1} \oplus \cdots \oplus (s_{n-1} \oplus e_{n-1}) \cdot h_{0,n-1} \\
r_1^e &= (s_n \oplus e_n) \cdot h_{0,0} \oplus (s_{n+1} \oplus e_{n+1}) \cdot h_{0,1} \oplus \cdots \oplus (s_{2n-1} \oplus e_{2n-1}) \cdot h_{0,n-1} \\
&\vdots \\
r_i^e &= (s_{ni} \oplus e_{ni}) \cdot h_{0,0} \oplus (s_{ni+1} \oplus e_{ni+1}) \cdot h_{0,1} \oplus \cdots \oplus \left(s_{n(i+1)-1} \oplus e_{n(i+1)-1}\right) \cdot h_{0,n-1}
\end{aligned} \tag{29}$$

We set d taps $1, i_1, \cdots, i_{d-1}$, whose corresponding polynomials are: $Q(X) = 1 + C_{i_1} x^{ni_1} + \cdots + C_{i_{d-1}} x^{ni_{d-1}}$. If $z_i^e = r_i^e \oplus r_{i-i_1}^e \oplus r_{i-r_2}^e \oplus \cdots \oplus r_{i-i_{d-1}}^e, i \geq i_{d-1}$ is assumed, then:

$$\begin{aligned}
& z_i^e = r_i^e \oplus r_{i-i_1}^e \oplus r_{i-r_2}^e \oplus \cdots \oplus r_{i-i_{d-1}}^e \\
& = \left((s_{ni} \oplus e_{ni}) h_{0,0} \oplus (s_{ni+1} \oplus e_{ni+1}) h_{0,1} \oplus \cdots \oplus (s_{ni+n-1} \oplus e_{ninn-1}) h_{0,n-1} \right) \\
& \oplus \left(\left(s_{n(i-i_1)} \oplus e_{n(i-i_i)} \right) h_{0,0} \oplus \left(s_{n(i-i,i)+1} \oplus e_{n(i-i_1)+1} \right) h_{0,1} \oplus \cdots \right. \\
& \left. \oplus \left(s_{n(i-i_i)+n-1} \oplus e_{n(i-i)+n-1} \right) h_{0,n-1} \right) \oplus \cdots \\
& \oplus \left(\left(s_{n(i-i_{-1})} \oplus e_{n(i-i_{d-1})} \right) h_{0,0} \oplus \left(s_{n(i-i_{d-1})+1} \oplus e_{n(i-i_{d-1})+1} \right) h_{0,1} \oplus \cdots \right. \\
& \left. \oplus \left(s_{n(i-i_{d-1})+n-1} \oplus e_{n(i-i_{d-1})+n-1} \right) h_{0,n-1} \right) \\
& = \left(s_{ni} \oplus s_{n(i-i_i)} \oplus \cdots \oplus s_{n(i-i_{\alpha-1})} \right) \cdot h_{0,0} \\
& \oplus \left(e_{ni} \oplus e_{n(i-i_i)} \oplus \cdots \oplus e_{n(i-i_{d-1})} \right) \cdot h_{0,0} \\
& \oplus \left(s_{ni+1} \oplus s_{ni+1-ni_1} \oplus \cdots \oplus s_{ni+1-nn_{d-1}} \right) \cdot h_{0,1} \\
& \oplus \left(e_{ni+1} \oplus e_{ni+1-ni_1} \oplus \cdots \oplus e_{ni+1-n_{d-1}} \right) \cdot h_{0,1} \\
& \vdots \\
& \oplus \left(s_{n(i+1)-1} \oplus \cdots \oplus s_{n(i+1)-1-ni_{d-1}} \right) \cdot h_{0,n-1} \\
& \oplus \left(e_{n(i+1)-1} \oplus \cdots \oplus e_{n(i+1)-1-nn_{i-1}} \right) \cdot h_{0,n-1}
\end{aligned} \tag{30}$$

We assume that w is the weight of the dual codeword h_0, and δ is the bit error rate of the historical information channel. According to the different historical information of inter-frame scrambling code, the composite code blind identification analysis of channel coding and scrambling code is as follows, when Q(X)=A(X)P(X):

$$\begin{gathered} s_{ni} \oplus s_{n(i-i)} \oplus \cdots \oplus s_{n(i-i_{-1})} = 0 \\ s_{ni+1} \oplus s_{n(i-i)+1} \oplus \cdots \oplus s_{n(i-i_{-1})+1} = 0 \\ \vdots \\ s_{n(i+1)-1} \oplus s_{n(i-i)-ni_1} \oplus \cdots \oplus s_{n(i-i_{d-1}+1)-1} = 0 \end{gathered} \tag{31}$$

Then,

$$\begin{aligned} z_i^e &= \left(e_{ni} \oplus e_{ni-ni_1} \oplus \cdots \oplus e_{ni-ni_{d-1}}\right) \cdot h_{0,0} \\ &\oplus \left(e_{ni+1} \oplus e_{ni+1-ni_1} \oplus \cdots \oplus e_{ni+1-ni_{d-1}}\right) \cdot h_{0,1} \oplus \cdots \\ &\oplus \left(e_{n(i+1)-1} \oplus \cdots \oplus e_{n(i+1)-1-ni_{d-1}}\right) \cdot h_{0,n-1} \end{aligned} \tag{32}$$

Then, the probability of $z_i^e = 1$ is:

$$\begin{aligned} Pr\left(z_i^e = 1\right) &= \sum_{l=1,3,\cdots}^{wd} C_{wd}^l \delta^l (1-\delta)^{wd-l} \\ &= \sum_{l=1,3,\cdots}^{wd} C_{wd}^l \left[\frac{1}{2} - \left(\frac{1}{2} - \delta\right)\right]^l \left[\frac{1}{2} + \left(\frac{1}{2} - \delta\right)\right]^{wd-l} \\ &= \frac{1}{2}\left[1 - 2\left(\frac{1}{2} - \delta\right)^{wd}\right] \\ &= \frac{1}{2} - \left(\frac{1}{2} - \delta\right)^{wd} \end{aligned} \tag{33}$$

If we assume that the length of the received sequence is N, the cumulant Z^e is Gaussian:

$$\begin{aligned} Z^e &= \sum_{t=ii_{j-1}}^{N-1} (-1)^{z=i} = \sum_{t=i}^{N-1} \left(1 - 2 \cdot z_{t-1}^e\right) \\ &= \sum_{t=i_{i-1}}^{N-1} 1 - 2 \cdot \sum_{t=i_{d-1}}^{N-1} z_t^e \\ &= \left(N - i_{d-1}\right) - 2 \sum_{t=i_{d-1}}^{N-1} z_t^e \end{aligned} \tag{34}$$

Its mean and variance are:

$$\mu = E\left[Z^e\right] = E\left[\left(N - i_{d-1}\right) - 2\sum_{t=i_{d-1}}^{N-1} z_t^e\right]$$
$$= \left(N - i_{d-1}\right) - 2\sum_{t=i-1}^{N-1} E\left[z_t^e\right]$$
$$= \left(N - i_{d-1}\right) - 2\left(N - i_{d-1}\right)\left[\left(1/2\right)\left(1 - 2\left(\frac{1}{2} - \delta\right)^{wd}\right)\right]$$
$$= \left(N - i_{d-1}\right)\left(\frac{1}{2} - \delta\right)^{wd} \tag{35}$$

$$\sigma^2 = \left(N - i_{d-1}\right)\left[1 + wd\left(\left(\frac{1}{2} - \delta\right)^2 - \left(\frac{1}{2} - \delta\right)^{2wd}\right)\right] \tag{36}$$

When $Q(X) \neq A(X)P(X)$, due to $Pr\left(s_i = 1\right) = 1/2$, there are:

$$Pr\left(s_{ni}{}' = 1\right) = Pr\left(s_{ni} \oplus s_{n(i-i)} \oplus \cdots \oplus s_{n(i-i-i-1)}\right) = 1\Big) = \frac{1}{2} \tag{37}$$

According to the literature, there are:

$$Pr\left(z_i^e = 1\right) = \frac{1}{2} \tag{38}$$

At this time, the cumulant Z^e obeys a Gaussian distribution with a mean of 0 and a variance of $N - i_{d-1}$.

If it is assumed that the detection threshold is T, when $\left|Z^e\right| \geq T$, the corresponding $Q(X)$ is a multiple of $P(X)$.When $\left|Z^e\right| < T$, the corresponding $Q(X)$ is not a multiple of $P(X)$.

The Y matrix constructed from the received sequence y_t^e is:

$$\mathrm{Y} = \begin{bmatrix} y_1^e \\ y_2^e \\ \vdots \\ y_m^e \end{bmatrix} = \begin{bmatrix} y_1^e & y_2^e & \cdots & y_n^e \\ y_{n+1}^e & y_{n+2}^e & \cdots & y_{2^*n}^e \\ \vdots & \vdots & \vdots & \vdots \\ y_{m^*n+1}^e & y_{m^*n+2}^e & \cdots & y_{(m+1)^*n}^e \end{bmatrix} \tag{39}$$

$y_m^e = \left(y_{(m-1)^*n+1}^e, y_{(m-1)^*n+2}^e, \cdots, y_{m^*n}^e\right)$, the Y matrix is extracted row by row:

$$\begin{aligned} C_i' &= y_i^e \oplus y_{i+i_1}^e \oplus \cdots \oplus y_{i+i_{d-1}}^e \\ &= \left(C_i \oplus m_i \oplus e_i\right) \oplus \left(C_{i-i_i} \oplus m_{i-i_1} \oplus e_{i-i_1}\right) \oplus \cdots \\ &\oplus \left(C_{i-i_{d-1}} \oplus m_{i-i_{d-1}} \oplus e_{i-i_{d-1}}\right) \\ &= \left(C_i \oplus C_{i-i_1} \oplus \cdots \oplus C_{i-i_{d-1}}\right) \oplus \left(m_i \oplus m_{i-i_1} \oplus \cdots \oplus m_{i-i_{i-1}}\right) \\ &\oplus \left(e_i \oplus e_{i-i_1} \oplus \cdots \oplus e_{i-i_{d-1}}\right) \end{aligned} \tag{40}$$

Similarly, when Q(X)=A(X)P(X), there are:

$$\begin{aligned} C_i' &= y_i^e \oplus y_{i_1}^e \oplus \cdots \oplus y_{i_{d-1}}^e \\ &= \left(C_i \oplus C_{i-i_1} \oplus \cdots \oplus C_{i-i_{-1}}\right) \oplus \left(e_i \oplus e_{i-i_1} \oplus \cdots \oplus e_{i-i_{d-1}}\right) \end{aligned} \tag{41}$$

Because $CH^T \equiv 0$, there is:

$$\begin{aligned} z_i^e &= max\left(C_i' H^T\right) \\ &= max\left(\left(\left(C_i \oplus C_{i-i_1} \oplus \cdots \oplus C_{i-i_{d-1}}\right) \oplus \left(e_i \oplus e_{i-i_i} \oplus \cdots \oplus e_{i-i_{-1}}\right)\right) H^T\right) \\ &= max\left(\left(e_i \oplus e_{i-i_i} \oplus \cdots \oplus e_{i-i_{\alpha-1}}\right) H^T\right) \\ &= max\left(\left[e_{ni} \oplus \cdots \oplus e_{ni-n_{d-1}} e_{ni+1} \oplus \cdots \oplus e_{ni+1-nn_{d-1}} \cdots e_{n(i+1)-1} \oplus \cdots \oplus e_{n(i+1)-1-nn_{\alpha-1}}\right] H^T\right) \end{aligned} \tag{42}$$

The code weight of the check matrix row vector is $w_j, 1 \le j \le n-k$, and the row vector of the check matrix is expressed as:

$$H = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_{(n-k)} \end{bmatrix} \tag{43}$$

For the j-th dual codeword, there is:

$$\begin{aligned} &\left[e_{nt} \oplus \cdots \oplus e_{nt-ni_{\alpha-1}} e_{nt+1} \oplus \cdots \oplus e_{nt+1-ni_{d-1}} \cdots e_{n(t+1)-1} \oplus \cdots \oplus e_{n(t+1)-1-ni_{\alpha-1}}\right] h_j^T \\ &= \left(e_{nt} \oplus e_{nt-nn_i} \oplus \cdots \oplus e_{nt-n_{d-1}}\right) \cdot h_{j,0} \\ &\oplus \left(e_{nt+1} \oplus e_{nt+1-nn_1} \oplus \cdots \oplus e_{nt+1-nn_{d-1}}\right) \cdot h_{j,1} \oplus \cdots \\ &\oplus \left(e_{n(t+1)-1} \oplus \cdots \oplus e_{n(t+1)-1-n_{d-1}}\right) \cdot h_{j,n-1} \end{aligned} \tag{44}$$

When the code weights of the dual codewords are equal and both are w:

$$z_i^e = \left[e_{ni} \oplus \cdots \oplus e_{ni-ni_{d-1}} e_{ni+1} \oplus \cdots \oplus e_{ni+1-n_{d-1}} \cdots e_{n(i+1)-1} \oplus \cdots \oplus e_{n(i+1)-1-ni_{\alpha-1}}\right] \mathrm{h}_j^T \tag{45}$$

Then the probability of $z_i^e = 1$ is:

$$\begin{aligned} Pr\left(z_i^e = 1\right) &= \sum_{l=1,3,\cdots}^{wd} C_{wd}^l \delta^l (1-\delta)^{wd-l} \\ &= \frac{1}{2} - \left(\frac{1}{2} - \delta\right)^{wd} \end{aligned} \tag{46}$$

When the code weights of the dual codewords are not equal, the dual codeword in the parity check matrix has the highest probability of z_i^e being 1 is h_{max}, and its code weight is w_{max}, then there are:

$$\begin{aligned} z_i^e &= max\left(\mathrm{C}_i \mathrm{H}^T\right) \\ &= \left[e_{ni} \oplus \cdots \oplus e_{ni-nin_{\alpha-1}} e_{ni+1} \oplus \cdots \oplus e_{ni+1-nn_{\alpha-1}} \cdots e_{n(i+1)-1} \oplus \cdots \oplus e_{n(i+1)-1-n_{\alpha-1}}\right] \mathrm{h}_{max} \end{aligned} \tag{47}$$

The cumulative amount Z^e is:

$$\begin{aligned} Z^e &= \sum_{t=i_{d-1}}^{N-1} (-1)^{z_i^e} = \sum_{t=i_{d-1}}^{N-1} \left(1 - 2z_t^e\right) \\ &= \sum_{t=i_{d-1}}^{N-1} 1 - 2\sum_{t=i_{d-1}}^{N-1} z_t^e \\ &= \left(N - i_{d-1}\right) - 2\sum_{t=i_{d-1}}^{N-1} z_t^e \end{aligned} \tag{48}$$

Its mean and variance are:

$$\begin{aligned} \mu = E\left[Z^e\right] &= E\left[\left(N - i_{d-1}\right) - 2\sum_{t=i_{d-1}}^{N-1} z_t^e\right] \\ &= \left(N - i_{d-1}\right) - 2\sum_{t=i_{d-1}}^{N-1} E\left[z_t^e\right] \\ &= \left(N - i_{d-1}\right) - 2\left(N - i_{d-1}\right)\left[(1/2)\left(1 - 2\left(\frac{1}{2} - \delta\right)^{w_{max} d}\right]\right] \\ &= \left(N - i_{d-1}\right)\left(\frac{1}{2} - \delta\right)^{w_{max} d}\right] \end{aligned} \tag{49}$$

$$\sigma^2 = \left(N - i_{d-1}\right)\left[1 + w_{max} d\left(\left(\frac{1}{2} - \delta\right)^2 - \left(\frac{1}{2} - \delta\right)^{2 w_{max} d}\right)\right] \tag{50}$$

From formula (49) and formula (50), it can be seen that the distribution of the cumulant Z^e is fixed. Whether the current Q(X) is a multiple of P(X) can be detected by detecting the threshold T.

3 ANALYSIS OF THE EVOLUTION OF MODERN CHINESE HISTORY BASED ON DATA MINING

The aggregation of historical archives resources has certain requirements for the organization of resources, and it is necessary to organize resources into a knowledge organization system under a specific aggregation framework. On the basis of clarifying the requirements, principles and processes of historical archives resource aggregation, this paper proposes a historical archives resource aggregation framework based on linked data. Moreover, this paper attempts to form a historical archive resource aggregation framework from three levels: data layer, linked data release aggregation layer, and linked data service application layer. Its functional structure is shown in Figure 9.

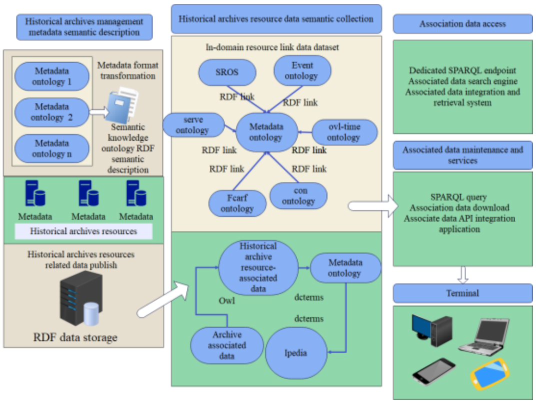


Figure 9: Framework of historical archive resource aggregation based on linked data.

The hierarchical syntax is generally adopted in the description of the entity relationship of the historical archive resources, and the description is carried out hierarchically. The first is to divide the object it faces - historical archives resources into four levels: description, organization, publication and application. Then, for the syntactic description of historical archive resources, the linked data standard is generally adopted to publish data. Then, data access is carried out with a systematic and unified data access system, the goal is to better promote the syntactic interoperability between heterogeneous data, and achieve the application results that the data is reasonable, satisfactory, and expected, as shown in Figure 10.

This paper verifies the effect of the model proposed in this paper, and analyzes the evolution process of modern Chinese history. Moreover, this paper verifies the effect of the model through multiple exercises, and obtains the test results shown in Table 1.

The experimental results in Table 1 verify that the evolution method of modern Chinese history based on data mining proposed in this paper can play an important role in historical data processing.

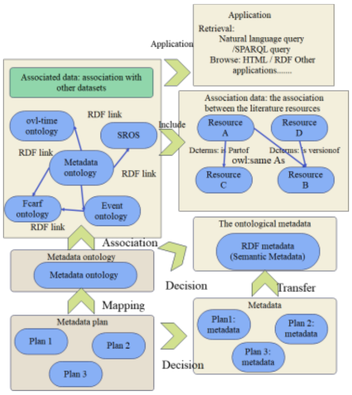


Figure 10: Description of the entity relationship of historical archive resources.

NO.	*Historical evolution*	*NO.*	*Historical evolution*	*NO.*	*Historical evolution*
1	*86.09*	*25*	*83.64*	*49*	*85.59*
2	*81.19*	*26*	*85.81*	*50*	*85.00*
3	*79.56*	*27*	*86.43*	*51*	*86.31*
4	*80.51*	*28*	*79.41*	*52*	*79.13*
5	*81.69*	*29*	*83.26*	*53*	*84.56*
6	*79.43*	*30*	*80.12*	*54*	*84.00*
7	*82.77*	*31*	*81.32*	*55*	*85.56*
8	*80.52*	*32*	*83.87*	*56*	*85.03*
9	*81.11*	*33*	*81.46*	*57*	*83.92*
10	*82.10*	*34*	*87.85*	*58*	*84.40*
11	*83.86*	*35*	*82.73*	*59*	*81.79*
12	*83.85*	*36*	*87.55*	*60*	*82.12*
13	*79.59*	*37*	*84.80*	*61*	*80.41*
14	*84.58*	*38*	*80.27*	*62*	*82.65*
15	*84.49*	*39*	*80.88*	*63*	*82.79*

16	*80.91*	*40*	*86.92*	*64*	*87.61*
17	*87.23*	*41*	*87.00*	*65*	*87.53*
18	*85.83*	*42*	*79.47*	*66*	*88.00*
19	*79.46*	*43*	*85.09*	*67*	*84.24*
20	*81.54*	*44*	*83.00*	*68*	*82.38*
21	*87.98*	*45*	*87.59*	*69*	*82.68*
22	*85.27*	*46*	*86.92*	*70*	*84.90*
23	*80.06*	*47*	*85.70*	*71*	*83.65*
24	*83.38*	*48*	*79.14*	*72*	*80.32*

Table 1: Analysis of the evolution of modern Chinese history based on data mining.

4 CONCLUSION

Building a cultural relic knowledge map for creative design is an important requirement for designers to use cultural relics for creative design. Traditional cultural relics information management systems use relational databases to store cultural relics knowledge. When designers search for ideas, it is difficult to obtain satisfactory results. Because the traditional cultural relics information retrieval methods mostly use keyword matching to retrieve the database, the system cannot understand the designer's retrieval needs at the semantic level, so the retrieval results are not so satisfactory. The traditional database lacks the expression specification of cultural relic knowledge and the derivation ability based on knowledge, so the ability to expand and retrieve knowledge is relatively weak. This paper combines the data mining algorithm to analyze the evolution of modern Chinese history, and improves the analysis effect of the evolution of modern history through the intelligent model. The experimental results verify that the evolution method of modern Chinese history based on data mining proposed in this paper can play an important role in historical data processing.

Huaili Hu, https://orcid.org/0009-0005-2368-4382

REFERENCES

[1] Andron, S.: Selling Streetness as Experience: The Role of Street Art Tours in Branding the Creative City, The Sociological Review, 66(5), 2018, 1036-1057. https://doi.org/10.1177/0038026118771293

[2] Badir, Y. F.; Frank, B.; Bogers, M.: Employee-Level Open Innovation in Emerging Markets: Linking Internal, External, and Managerial Resources, Journal of the Academy of Marketing Science, 48(5), 2020, 891-913. https://doi.org/10.1007/s11747-019-00674-6

[3] Bhansing, P. V.; Hitters, E.; Wijngaarden, Y.: Passion inspires: Motivations of Creative Entrepreneurs in Creative Business Centres in the Netherlands, The Journal of

Entrepreneurship, 27(1), 2018, 1-24. https://doi.org/10.1177/0971355717738589
[4] Bilan, Y.; Vasilyeva, T.; Kryklii, O.; Shilimbetova, G.: The Creative Industry as a Factor in the Development of the Economy: Dissemination of European Experience in the Countries with Economies in Transition, Creativity Studies, 12(1), 2019, 75-101. https://doi.org/10.3846/cs.2019.7453
[5] Campbell, P.; O'brien, D.; Taylor, M.: Cultural Engagement and the Economic Performance of the Cultural and Creative Industries: An Occupational Critique, Sociology, 53(2), 2019, 347-367. https://doi.org/10.1177/0038038518772737
[6] Campuzano, M. V.: Force and inertia: A systematic Review of Women's Leadership in Male-Dominated Organizational Cultures in the United States, Human Resource Development Review, 18(4), 2019, 437-469. https://doi.org/10.1177/1534484319861169
[7] Charoensukmongkol, P.: The Efficacy of Cultural Intelligence for Adaptive Selling Behaviors in Cross-Cultural Selling: The Moderating Effect of Trait Mindfulness, Journal of Global Marketing, 33(3), 2020, 141-157. https://doi.org/10.1080/08911762.2019.1654586
[8] Comunian, R.; England, L.: Creative and Cultural Work Without Filters: Covid-19 and Exposed Precarity in the Creative Economy, Cultural Trends, 29(2), 2020, 112-128. https://doi.org/10.1080/09548963.2020.1770577
[9] Eikhof, D. R.: COVID-19, Inclusion and Workforce Diversity in the Cultural Economy: What Now, what next?, Cultural Trends, 29(3), 2020, 234-250. https://doi.org/10.1080/09548963.2020.1802202
[10] Kalnitska, M. O.: Assessment of the Development State of Organizational and Cultural Resources of International Tourism Business, European Journal of Management Issues, 26(3-4), 2018, 71-81. https://doi.org/10.15421/191808
[11] Lehto, X. Y.; Lehto, M. R.: Vacation as a Public Health Resource: Toward a Wellness-Centered Tourism Design Approach, Journal of Hospitality & Tourism Research, 43(7), 2019, 935-960. https://doi.org/10.1177/1096348019849684
[12] Muafi, M.: A Nexus Among Strategic Orientation, social Network, Knowledge Sharing, Organizational Innovation, and MSMEs Performance, The Journal of Asian Finance, Economics and Business, 7(6), 2020, 327-338. https://doi.org/10.13106/jafeb.2020.vol7.no6.327
[13] Olalekan, R. M.; Omidiji, A. O.; Williams, E. A.; Christianah, M. B.; Modupe, O.: The Roles of all Tiers of Government and Development Partners in Environmental Conservation of Natural Resource: a Case Study in Nigeria, MOJ Ecology & Environmental Sciences, 4(3), 2019, 114-121. https://doi.org/10.15406/mojes.2019.04.00142
[14] Sabuhari, R.; Sudiro, A.; Irawanto, D.; Rahayu, M.: The Effects of Human Resource Flexibility, Employee Competency, Organizational Culture Adaptation and Job Satisfaction on Employee Performance, Management Science Letters, 10(8), 2020, 1775-1786. https://doi.org/10.5267/j.msl.2020.1.001
[15] Sasongko, T.; Rifa'i, M.; Sayekti, N. S.: The Development of the Creative Industries to Create a Competitive Advantage: Studies in Small Business Sector, Journal of Economic Development, Environment and People, 7(3), 2019, 14-23.
[16] Soelton, M.: How culture, Training Standard and Discipline on the Employee Performance Affect Hotel Management, European Research Studies, 21(4), 2018, 378-385. https://doi.org/10.35808/ersj/1128
[17] Stoermer, S.; Davies, S.; Froese, F. J.: The Influence of Expatriate Cultural Intelligence on Organizational Embeddedness and Knowledge Sharing: The Moderating Effects of Host Country Context, Journal of International Business Studies, 52(3), 2021, 432-453. https://doi.org/10.1057/s41267-020-00349-3