International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-16

https://doi.org/10.70917/ijcisim-2026-0162
Article

Modeling and Research on the Relationship between Japanese Monster Culture and Folk Beliefs in a Big Data Environment

Hong Zhang *

School of Foreign Languages, Sanda University, Shanghai, 201209, China; zhanghonfg4807@163.com

Abstract: This paper aims to explore the intrinsic relationship between Japanese yokai culture and folk beliefs in order to gain a deeper understanding of cultural contexts and the foundations of belief systems. First, the collected data was preprocessed using the Jieba word segmentation tool and stopword removal methods. Textual data was converted into computable structured data using a vector space model. The TF-IDF algorithm was employed to extract text features and perform dimensionality reduction, resulting in features with simple and clear vector dimensions. Next, a multidimensional network of Japanese yokai culture and folk beliefs is constructed, and social network analysis is applied to investigate the strength of feature associations between Japanese yokai culture and folk beliefs. Association rule algorithms are then introduced to further explore the association patterns between Japanese yokai culture and folk beliefs, and the obtained association relationships are visualized. The study shows that the average shortest path in the relationship network between Japanese yokai culture and folk beliefs is only 1.05. This network is influenced by multiple factors such as yokai types, regional distribution characteristics, and historical periods, indicating that the network has high interaction efficiency, enabling the rapid dissemination of Japanese yokai culture. It is not easily disrupted by external factors, making the relationship network between Japanese yokai culture and folk beliefs more adaptable over time and conducive to cultural inheritance.

Keywords: Vector Space Model; TF-IDF Algorithm; Social Network Analysis Method; Association Rule Algorithm; Japanese Monster Culture

1. Introduction

Folk beliefs are a set of concepts, behavioral habits, and corresponding ritual systems related to the worship of deities that have spontaneously emerged among the populace over the course of long-term historical development. They represent the spontaneous veneration and respect of the populace for spiritual entities endowed with supernatural powers [1-3]. This includes the transmission of primitive religions among the populace, the penetration of man-made religions into folk beliefs, widespread popular superstitions, and general popular superstitions. Japanese yokai culture is a reflection of animism, primarily manifested through soul worship, and as such, folk beliefs are increasingly attracting attention [4-5].

In the context of the global digital revolution, the dissemination of Japanese yokai culture and folk beliefs is no longer confined to traditional modes, gradually forming a multi-dimensional dissemination system that transitions from offline exhibitions to online preservation, from one-way dissemination to two-way interaction, and from closed-door inheritance to open-door sharing [6-7]. As audiences watch, like, comment on, and share content related to Japanese yokai culture and folk beliefs, the dissemination of these cultural elements has become more diverse and socialized [8-9]. Traditional yokai in folk beliefs constitute an important part of Japanese religious beliefs, and their historical origins and symbolic meanings warrant further exploration.

Currently, there are few reports on the digital dissemination of Japanese yokai culture and folk beliefs. However, relevant studies indicate that digital dissemination, with its interactive, immersive, and decentralized technical characteristics, has reshaped the dissemination pathways of traditional culture, driving audiences to transition from passive recipients to active participants [10]. For example, the literature [11] uses the Shu brocade weaving technique as a case study to analyze the effectiveness of cultural digital dissemination in a big data environment. The research results indicate that the effectiveness of cultural dissemination has shifted from a traditional technology-oriented approach to one centered on cultural content, authenticity, and integrity, thereby promoting the sustainable development of culture. Reference [12] uses cultural heritage sites as an example to demonstrate the effectiveness of virtual technology in the digital dissemination of cultural heritage through the integration of cultural characteristics into cyberspace using virtual reality technology. This is proven from three aspects: interactivity, spatial perception, and cultural value presentation. Reference [13] points out that computer 3D technology can fully and dynamically display intangible cultural heritage, improving its reproducibility and providing useful technical support for the training and guidance of cultural heritage bearers. Reference [14] In order to promote the digital dissemination and protection of China's intangible cultural heritage, a digital dissemination platform for intangible cultural heritage was established. The platform achieved a coverage rate of 92.5% and a popularization rate of 89.8% in the fields of folklore and traditional craftsmanship knowledge, thereby facilitating the widespread dissemination and protection of intangible cultural heritage in the era of big data.

The rapid development of big data technology has provided new opportunities for the dissemination of culture, enabling it to break through the limitations of traditional dissemination models and enter the public eye in more flexible and diverse ways, gaining popularity among a wider audience. However, further research is needed on the digital dissemination of Japanese yokai culture and folk beliefs.

After collecting and processing basic text data, this paper combines network characteristics and centrality analysis methods from social network analysis, the Apriori association rule algorithm, and visualization image processing technology based on the R language application software and Gephi tools. It delves deeply into the intrinsic connection between Japanese yokai culture and folk beliefs, constructs quantitative models for both, reveals the network characteristics of Japanese yokai culture and folk beliefs, and confirms the role and value of Japanese yokai culture and folk beliefs in cultural inheritance.

2. Modeling the Relationship between Japanese yokai culture and folk beliefs

2.1. Text Data Feature Analysis

In modeling the relationship between Japanese yokai culture and folk beliefs, a large amount of data has been accumulated, which also possesses the “5Vs” characteristics of big data proposed by IBM, namely volume, variety, velocity, veracity, and low value density.

2.2. Text Data Preprocessing Methods

2.2.1. Chinese Word Segmentation of Text Data

To achieve Chinese word segmentation, it is necessary to combine algorithms with analytical tools. Commonly used tools include the Chinese word segmentation tools Jieba and THULAC, NEUSP, Pangu Word Segmentation, Paoding Word Segmentation, and the automatic word segmenter NLPWi developed by Microsoft Research.

This paper selects the Jieba word segmentation tool, whose package is compatible with Python language coding, supports user-defined dictionaries, and provides multiple word segmentation engine modes. It offers superior word segmentation speed and effectiveness, enabling efficient implementation of Chinese word segmentation, word frequency statistics, stopword processing, and more. The preprocessed text data can be directly utilized for research into the relationship and visualization of connections between Japanese yokai culture and folk beliefs.

2.2.2. Customizing the Removal of Stop Words

Stop words refer to words that frequently appear in text but have no functional meaning. Due to the relationship between Japanese yokai culture and folk beliefs, textual records often contain non-standard terminology. After word segmentation, custom stop words must be input to filter out and delete words that are meaningless or lack value, thereby avoiding dimensionality explosion, reducing noise, and improving the efficiency and effectiveness of text mining. Stop words primarily fall into the following categories:

(1) Words that serve functional purposes in Chinese, such as expressing emotions or smoothing sentence structure, but have no actual meaning for text topic analysis.

(2) Specific vocabulary that does not contribute value to text mining and can be filtered out to reduce dimensionality.

(3) The third category includes punctuation marks and irrelevant data that need to be removed to avoid interference with vector representation, weight calculation, topic mining, text clustering, etc.

2.2.3. Text Vector Space Model

After preprocessing, the text remains as textual data and needs to be converted into structured data that is recognizable and computable by computers through a vector space model. The Vector Space Model (VSM) [15] is widely applied in fields such as text classification, information retrieval, and predictive analysis. The fundamental concept of VSM is to represent text as vectors, where the vector dimensions correspond to the weights of feature words in the document collection. This representation is referred to as a bag-of-words. After preprocessing the text on the relationship between Japanese yokai culture and folk beliefs, the resulting terms form documents, which can be represented as:

$$D_i = (w_1, w_2, \cdots, w_k, \cdots, w_n) \tag{1}$$

Among them, D_i is the i th document in the text set D, and $1 \le i \le |D|, D_i \subseteq D$; w_k represents the k th feature item, and $1 \le k \le n$.

By assigning a corresponding weight t_k to each feature term w_k based on its importance in the document, we can represent the terms and weights in the entire text corpus:

$$D = (w_1, t_1; w_2, t_2; \cdots; w_k, t_k; \cdots; w_n, t_n) \tag{2}$$

In this context, t_k denotes the weight of the feature term w_k, where $1 \le k \le n$; D denotes the entire document collection. In the Vector Space Model (VSM), the following conventions apply:

(1) Each feature term $w_k (1 \le k \le n)$ is distinct, i.e., there are no duplicates.

(2) The feature terms w_k have no order relationship, i.e., the internal structure of the document is not considered.

Under the above two conventions, feature terms can be represented as an n -dimensional coordinate, where the weight $(t_1, t_2, \cdots, t_k, \cdots, t_n)$ of each feature term is the corresponding coordinate value. A text can then be represented as a vector in an n -dimensional space:

$$D_i = (w_{i1}, t_{i1}; w_{i2}, t_{i2}; \cdots; w_{ik}, t_{ik}; \cdots; w_{in}, t_{in}) \tag{3}$$

$\theta (1 \le \theta \le \pi)$ is the angle between text vectors D_1 and D_2, thereby converting textual data on the relationship between Japanese yokai culture and folk beliefs into a vector space representation, which is then transformed into structured data that can be recognized and calculated by computers. This study uses Python software to implement the conversion of the text vector space model, facilitating research on the relationship between Japanese yokai culture and folk beliefs, as well as the visualization of these relationships.

2.2.4. Calculation of text feature weights

Different word frequencies contribute differently to the text, and a large number of terms can lead to complex feature vector dimensions. Terms are categorized into high-frequency and low-frequency words based on their frequency of appearance in documents. Each term is assigned a weight based on its frequency, and terms with low weights are filtered out to achieve feature extraction and dimensionality reduction. The TF-IDF algorithm is commonly used for this purpose.

(1) Term Frequency (TF) Method

Term frequency (TF) represents the frequency of occurrence of risk term i in construction risk record documents. A term with a high frequency of occurrence in a document is considered a high-frequency term, indicating a greater contribution to text analysis, while a term with a low frequency of occurrence is considered a low-frequency term, with lesser analytical value. Therefore, a term frequency filter can be set to remove low-frequency terms, achieving dimension reduction and improving text mining effectiveness. Converting term frequency into weights for calculation:

$$tf_i = \frac{n_{i,j}}{\sum_K n_{k,j}} \tag{4}$$

Among these, $n_{i,j}$ represents the frequency of occurrence of risk term i in construction risk record document d_j; the denominator denotes the total number of all feature terms in construction risk record document d_j.

(2) Inverse Document Frequency (IDF) Method

Inverse Document Frequency (IDF) is used to measure the contribution or importance of a word to the entire document. Contrary to term frequency, the fewer times a word appears in a document, the higher its inverse document frequency, indicating its uniqueness, and vice versa. Its weight calculation:

$$idf_i = \log \frac{|D|}{1+|\{j : t_i \in d_i\}|} \tag{5}$$

Among them, $|D|$ represents the total number of documents; $|\{j : t_i \in d_i\}|$ represents the number of documents containing the descriptive term t_i. To avoid a denominator of 0, 1 is added for processing.

(3) Term Frequency-Inverse Document Frequency (TF-IDF) Algorithm

Term Frequency-Inverse Document Frequency (TF-IDF) [16] is a highly applicable weighted algorithm in text mining, combining the advantages of term frequency and inverse document frequency to measure the importance of terms, offering higher precision than the TextRank algorithm. The specific formula for TF-IDF is:

$$(tf_i - idf_i)_{i,j} = tf \times idf_i = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log \frac{|D|}{1+|\{j : t_i \in d_i\}|} \tag{6}$$

The higher the TF value of a word, the higher its frequency and contribution to text analysis, so it is assigned a higher weight and extracted as a high-frequency word. When the word appears more frequently in other documents, its uniqueness and distinctiveness to the document decrease, resulting in a lower IDF value. Its importance also decreases as the word appears more frequently in other documents within the entire document collection. By filtering out and removing low-frequency, frequently occurring words using the TF-IDF algorithm, the weighting impact of redundant words can be reduced, resulting in better theme extraction effects for the retained feature words.

2.3. Social Network Analysis Theory

2.3.1. Network Density and Correlation Analysis

The network density is expressed as the ratio of the actual number of node connections in the overall network to the maximum possible number of connections. Let the number of nodes in the overall network be n, the actual number of connections in the overall network be m, and the network density D_n be expressed as.

$$D_n = \frac{m}{n \times (n-1)} \tag{7}$$

In social network analysis [17], connectivity reflects the stability of the overall network. If the connections between nodes in the overall network can link all nodes into a network system, with any two nodes connected by a direct or indirect path, then the overall network is considered stable and has high connectivity. If multiple connections in the overall network pass through a fixed node, then the overall network is considered to be highly dependent on that node. If this node is removed, the network system may collapse, so the overall network is considered unstable and has low connectivity. The metric for measuring connectivity in social network analysis is the connectivity coefficient C. Let the number of nodes in the overall network be n, and the number of unreachable pairs of nodes in the network be v. The connectivity coefficient C_n is expressed as:

$$C_n = 1 - \frac{v}{n \times (n-1)/2} \tag{8}$$

The network degree reflects the extent to which nodes in the overall network are asymmetrically reachable, describing the dominance of each node in the overall network. Let the number of symmetrically reachable pairs in the overall network be K, and the maximum possible number of symmetrically reachable pairs be $\max(K)$. The degree H is expressed as:

$$H = 1 - \frac{K}{\max(K)} \tag{9}$$

2.3.2. Centrality Analysis

Centrality is the focus of social network analysis of the relationship between Japanese yokai culture and folk beliefs. Social network centrality analysis has three core indicators: degree centrality, betweenness centrality, and closeness centrality, which respectively express the connectivity, accessibility, and intermediary nature of nodes.

(1) Degree centrality (hereinafter referred to as D): This refers to the number of direct links to a research object (node). The formula for calculating D is:

$$D = \frac{x}{(n-1)} \tag{10}$$

In the formula, n represents the number of nodes, and x represents the number of connections between a given node and other nodes. The larger D is, the better the connectivity of the node.

(2) Centrality (hereinafter referred to as C): This refers to the average length of the shortest distance between each research object (node). The formula for calculating C is as follows:

$$C \frac{(n-1)}{\sum_{j=1}^{n} d_{ij}} \tag{11}$$

In the formula, d_{ij} represents the shortest distance between research objects (nodes) i and j.

(3) Proximity to the center (hereinafter referred to as B): This measures the number of times a research object (node) is traversed on the shortest path between other research objects (nodes). The formula for calculating B is as follows:

$$B = \frac{2\sum_{a}^{n}\sum_{b}^{n} g_{ab}(i) / g_{ab}}{n^2 - 3n + 2} \tag{12}$$

In the formula, g_{ab} is the shortest path between the study object (node) a and b, and $g_{ab}(i)$ is the number of times i is traversed on the shortest path between a and b, where $a \neq b, a < b$.

In the network structure of the relationship between Japanese yokai culture and folk beliefs, betweenness centrality, closeness centrality, and degree centrality are used to calculate the convenience, accessibility, and intermediary nature of network nodes from three perspectives. However, any centrality measure can only reflect one aspect of the network structure and cannot reflect the overall characteristics of the network space structure.

Considering that degree centrality, betweenness centrality, and closeness centrality are representative and exhibit a certain degree of correlation, satisfying the computational conditions for factor analysis, the weights of the three types of centrality can be calculated using factor analysis in SPSS. Due to the significant differences in the value ranges of degree centrality, betweenness centrality, and closeness centrality, the three indicators are first normalized to obtain standardized values.

The calculation formula for geographic centrality (hereinafter referred to as G) is as follows:

$$G = w_D \times \frac{x_{Di} - \overline{x_D}}{s_D} + w_C \times \frac{x_{Ci} - \overline{x_C}}{s_C} + w_B \times \frac{x_{Bi} - \overline{x_B}}{s_B} \tag{13}$$

$$\overline{x_D} = \frac{\sum_{i=1}^{n} x_{Di}}{n} \tag{14}$$

$$\overline{x_C} = \frac{\sum_{i=1}^{n} x_{Ci}}{n} \tag{15}$$

$$\overline{x_B} = \frac{\sum_{i=1}^{n} x_{Bi}}{n} \tag{16}$$

$$s_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{Di} - \overline{x_D})^2} \tag{17}$$

$$s_C = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{Ci} - \overline{x_C})^2} \tag{18}$$

$$s_B = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{Bi} - \overline{x_B})^2} \tag{19}$$

In the formula, w_D, w_C, w_B represent the weights of point centrality, intermediary centrality, and proximity centrality, respectively, which are calculated using the entropy method. $\overline{x_D}, \overline{x_C}, \overline{x_B}$ represent the means of point centrality, intermediary centrality, and proximity centrality, respectively. s_D, s_C, s_B represent the standard deviations of degree centrality, betweenness centrality, and closeness centrality, respectively.

2.3.3. Cohesive subgroup analysis

A cohesive subgroup [18] refers to a subset (or object) within a specified network that exhibits strong, dense, active, and interconnected relationships among its members.

Cohesive subgroups can be defined and analyzed using different methods based on theoretical principles and computational techniques.

(1) Clique: In an undirected network graph, a “clique” refers to the largest complete subgraph containing at least three nodes. This concept has three characteristics: (a) a clique must include at least three points; (b) a clique is complete, meaning that any two points within the clique are directly connected; (c) a clique is “maximum,” meaning that adding any additional point to this subgraph would alter its “complete” nature.

(2) n-clique: In a total graph, if the distance (i.e., the shortest path) between any two points in the subgraph is no more than n in the total graph. From the formal perspective of graph theory, let d_{ij} represent the distance between two points and n in the total graph. Then, the mathematical definition of an n-clique is a subgraph with a set of points that satisfies the following conditions: $d_{ij} <= n(n_i, n_j \in N)$, and there is no point in the total graph whose distance from any point in the subgraph exceeds n. In this case, the subgraph is referred to as an n-clique.

(3) n-clan: refers to an n-clan that satisfies the condition that “the shortest distance between any two points does not exceed n.” It can be seen that all n-clans are n-clans.

(4) k-clusters. A k-cluster refers to a cohesive subgroup that satisfies the condition that “in a given subgroup, each point is directly connected to all other points except for k points.” That is, when the size of this cohesive subgroup is n, each point is directly connected to at least n-k points within the subgroup, meaning that the degree of each point is at least n-k.

(5) Cohesive subgroup density: The density of cohesive subgroups is primarily used to measure the extent of small-group phenomena in a large network. This is particularly useful in analyzing complex

social network issues. If large groups are loosely organized while core small groups exhibit high cohesion, this warrants global attention; similarly, if large groups contain numerous small groups with high cohesion, this also warrants attention, as it may lead to malicious competition among small groups.

2.3.4. Core-Periphery Structure Analysis

Core-periphery structural analysis aims to investigate which research subjects (nodes) occupy a central position and which occupy a peripheral position in the social network of Japanese yokai culture and folk beliefs. Depending on the type of relationship data (categorical data and ratio data), core-periphery structures take different forms.

2.4. Apriori Association Rule Algorithm

The Apriori algorithm [19] is the first association rule mining algorithm and the most classic one. It uses an iterative method of sequential search to identify relationships between item sets in a database and form rules. The article uses statistical analysis and data mining tools in the SPSS Modeler software to build Apriori association rule models. Considering the complex and disordered nature of the output rules, the results are categorized across various dimensions, and the rules are interpreted under different dimensions to identify corresponding strong association rules and pinpoint the key influencing factors in the relationship between Japanese yokai culture and folk beliefs.

The construction of the model is closely related to the parameter values of the algorithm. When measuring and evaluating association rules, support, confidence, and lift are the three key parameters of the Apriori algorithm.

(1) Support refers to the frequency with which the antecedent and consequent appear simultaneously in a dataset. For example, when studying the association between “monster type” (X) and “medium of transmission” (Y) in the characteristics of the relationship between Japanese monster culture and folk beliefs, the support of the two parameters is expressed as:

$$\sup(X \to Y) = \frac{X \cap Y}{N} \tag{20}$$

In the formula, sup is the support degree; $X \cap Y$ is the frequency of occurrence of items X, Y; N is the total frequency of occurrence of items.

(2) Confidence degree The probability that the latter item will occur after the former item occurs can also be explained by the conditional probability of the feature. For example, the confidence degree that “monster type” causes a change in “transmission medium” is expressed as:

$$conf(X \to Y) = \frac{\sup(X \to Y)}{\sup(X)} \tag{21}$$

In the formula, conf is the confidence level.

(3) The lift reflects the degree of association between the antecedent and consequent in the association rule. A lift greater than 1 and the higher it is, the higher the positive correlation. A lift of 1 indicates no correlation, i.e., they are independent of each other. The lift is expressed as:

$$lift(X \to Y) = \frac{conf(X \to Y)}{\sup(Y)} \tag{22}$$

2.5. Visual Analysis

Visualization refers to the use of computer image processing technology to concretely represent data results in the form of images. This article uses the statistical application software R and the Gephi tool to analyze the results of data mining using images, which allows for a more intuitive understanding of the distribution patterns of factors influencing the relationship between Japanese monster culture and folk beliefs, as well as the degree of correlation between these factors.

3. Modeling the Relationship between Culture and Belief: Mining Association Rules for Influencing Factors

3.1. Association Rule Mining Results

Based on the mining steps of the Apriori association rule algorithm, a total of 370 association rules were ultimately identified, including 45 two-item set rules, 150 three-item set rules, 120 four-item set rules, 48 five-item set rules, and 7 six-item set rules. Binary set rules primarily reflect the association between individual factors influencing the modeling of relationships, while multi-item set rules primarily reflect the association between frequent item sets of factors influencing the modeling of relationships between Japanese yokai culture and folk beliefs. Figure 1 shows the support, confidence, and lift of the 370 association rules. Each scatter point in the figure represents an association rule. The horizontal axis of the figure represents the support of the rule. The vertical axis of the figure represents the confidence of the rule. The shade of the scatter point color represents the lift of the rule, with colors ranging from light to dark indicating a change in lift from low to high.

Among these 370 association rules, 88% of the association rules are located in the area between 0.2 and 0.375 on the horizontal axis, which means that nearly 90% of the association rules have a support between 0.2 and 0.375.

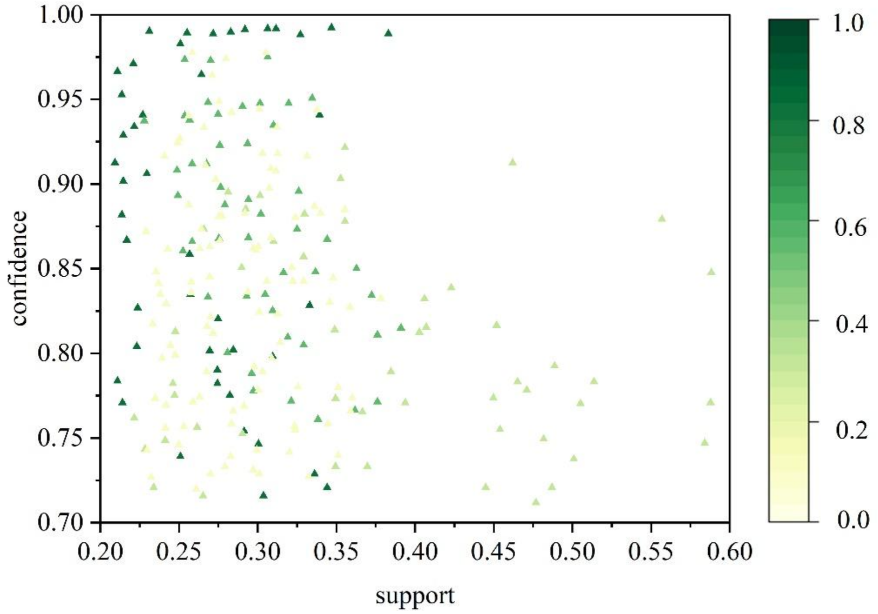

Figure 1. Support, confidence and ascension of association rules.

3.2. High Support Association Rules

Sort the 370 association rules according to their support degrees and select the top 40 association rules. The top 40 association rules are shown in Table 1 and are referred to as high-support association rules. Table 1 shows the antecedents and consequents of the top 40 high-support association rules, along with their corresponding support, confidence, and lift values. The support values range from 0.352 to 0.530, confidence values from 0.722 to 0.928, and lift values from 1.000 to 1.444. It can be observed that these 40 association rules are composed of six influencing factors: “Monster Type: Kappa (PRH3-1)”, “Cultural Transmission Medium (PRS2)”, “Regional Distribution Characteristics (EH1)”, “Public Participation Level (ES2)”, “Degree of Integration with Buddhist Beliefs (PRP6-1)”, and “Historical Period (PRP8-2)”. This indicates that these six influencing factors are closely interrelated and exhibit frequent associations, as detailed below: There are three rules with support values above 0.5. The factors of “monster type” and “cultural transmission medium” form the two rules with the highest support values among all rules, with a support value of 0.530. The next most frequently associated rule is composed of the factors of “monster type” and “regional distribution characteristics,” with a support value of 0.528. These three rules indicate that monster type has the most frequent associative relationships with cultural transmission medium and regional distribution characteristics. The six influencing factors—monster type, cultural transmission medium, degree of integration with Buddhist beliefs, regional distribution characteristics, and public participation—exhibit frequent associative relationships. They form associative rules through different combinations of factors, forming 40 high-support association rules, including two-item set rules and three-item set rules. This indicates that in the modeling of the relationship between Japanese monster culture and folk beliefs, these six factors frequently appear together and have strong associations. The appearance of a single factor or a set of factors triggers the appearance of other factors within this set. Historical periods and cultural transmission media have frequent associations. The historical period appears only once in the 40 high-support association rules, forming an association rule with a support degree as high as 0.442 with the cultural transmission medium factor. This indicates that the cultural transmission medium within the historical period is an important influencing factor in modeling the relationship between Japanese yokai culture and folk beliefs.

Table 1. Highlengthcorrelationrules.

Serial number	Pilotterm	Follow-onterm	Support	Confidence	Degreeofascension
1	PRH3-1	PRS2	0.530	0.928	1.000
2	PRS2	PRH3-1	0.528	0.902	1.062
3	EH1	PRH3-1	0.524	0.875	1.123
4	PRH3-1	EH1	0.498	0.849	1.185
5	PRP6-1	PRH3-1	0.498	0.823	1.246
6	PRH3-1	PRP6-1	0.462	0.801	1.297
7	ES2	PRH3-1	0.455	0.812	1.277
8	ES2	PRS2	0.456	0.822	1.256
9	EH1	PRS2	0.456	0.832	1.236
10	PRP8-2	PRS2	0.442	0.842	1.215
11	PRP6-1	EH1	0.442	0.852	1.214
12	EH1	PRP6-1	0.438	0.860	1.272
13	ES2	EH1	0.438	0.868	1.329
14	EH1	ES2	0.418	0.877	1.386
15	PRP6-1	ES2	0.418	0.885	1.443
16	ES2	PRP6-1	0.418	0.879	1.444
17	PRH3-1, EH1	PRS2	0.418	0.851	1.350
18	PRS2, EH1	PRH3-1	0.418	0.822	1.256
19	PRH3-1, PRS2	EH1	0.418	0.793	1.163
20	PRP6-1	PRS2	0.418	0.764	1.069
21	EH1, PRP6-1	PRH3-1	0.410	0.747	1.021
22	PRH3-1, PRP6-1	EH1	0.385	0.741	1.019
23	PRH3-1, EH1	PRP6-1	0.385	0.735	1.017
24	EH1, ES2	PRH3-1	0.385	0.728	1.015
25	PRH3-1, ES2	EH1	0.385	0.722	1.013
26	PRH3-1, EH1	ES2	0.385	0.725	1.018
27	PRH3-1, PRP6-1	PRS2	0.385	0.733	1.026
28	PRP6-1, PRS2	PRH3-1	0.370	0.742	1.035
29	PRH3-1, PRS2	PRP6-1	0.370	0.750	1.044
30	PRH3-1, ES2	PRS2	0.370	0.758	1.053
31	PRS2, ES2	PRH3-1	0.370	0.771	1.093
32	PRH3-1, PRS2	ES2	0.369	0.785	1.143
33	PRP6-1, ES2	PRH3-1	0.369	0.799	1.193
34	PRH3-1, PRP6-1	ES2	0.369	0.814	1.243
35	PRH3-1, ES2	PRP6-1	0.369	0.828	1.294
36	PRP6-1, ES2	EH1	0.369	0.844	1.282
37	PRP6-1, EH1	ES2	0.362	0.861	1.262
38	EH1, ES2	PRP6-1	0.362	0.877	1.241
39	EH1, ES2	PRS2	0.362	0.894	1.221
40	PRS2, ES2	EH1	0.352	0.910	1.200

3.3. High-confidence association rules

Sort all association rules by confidence level and extract the top 40 association rules with the highest confidence levels as high-confidence association rules. The high-confidence association rules are shown in Table 2. Table 2 displays the antecedent and consequent items of the high-confidence association rules, as well as the support, confidence, and lift values corresponding to each rule. The support values range from 0.202 to 0.351, the confidence values range from 0.952 to 1.000, and the lift values range from 1.353 to 2.621.

The high-confidence association rule results include association rules ranging from two-item sets to six-item sets, reflecting the influence of factor items and factor item sets in the antecedent items on the factors in the consequent items. The confidence level indicates the probability of the consequent item occurring when the antecedent item occurs, so the consequent items require special attention.

Among these 40 high-confidence association rules, the antecedents include “cultural transmission medium (PRS2),” “geographical distribution characteristics (EH1),” “public participation (ES2),” “degree of integration with Buddhist beliefs (PRP6-1),” “historical period (PRP8-2),” “gender attributes of monsters (EP4),” “degree of modern commercial development (EV2-3),” “Environmental Association (POS1),” “Social Structure (EV1-1),” “Cultural Flow (PRV1-3),” “Goodness/Evilness Attributes of Monsters (EP2),” and “Spatial-Temporal Distribution Characteristics (P0H2).” The subsequent items only include four influencing factors: “Degree of Integration with Buddhist Beliefs (PRP6-1),” “Geographical Distribution Characteristics (EH1),” “Yōkai Type: Kappa (PRH3-1),” and “Frequency of Official Document Records (EH2).” The rule confidence levels for these factors are all above 0.96, indicating that under the conditions where the preceding factors have already appeared, these four factors have an extremely high probability of occurring.

Table 2. High confidence association rules.

Serial number	Pilotterm	Follow-onterm	Support	Confidence	Degreeofascension
1	EP4	PRP6-1	0.331	1.000	1.785
2	EH2, EP4	PRP6-1	0.318	1.000	1.785
3	PRH3-1, EP4	PRP6-1	0.305	1.000	1.785
4	EP4, ES2	PRP6-1	0.292	1.000	1.785
5	EH2, EP4, ES2	PRP6-1	0.287	1.000	1.785
6	PRS2, EP4	PRP6-1	0.292	1.000	1.785
7	PRH3-1, EH2, EP4	PRP6-1	0.296	1.000	1.785
8	PRH3-1, EP4, ES2	PRP6-1	0.301	1.000	2.412
9	PRH3-1, EP4, ES2	PRP6-1	0.306	1.000	2.412
10	PRH3-1, EH2, EP4, ES2	PRP6-1	0.310	1.000	2.412
11	PRS2, EP4, ES2	PRP6-1	0.303	1.000	2.412
12	PRS2, EH2, EP4	PRP6-1	0.293	1.000	2.412
13	PRH3-1, PRS2, EP4, ES2	PRP6-1	0.283	1.000	1.329
14	PRH3-1, PRS2, EP4, EH2	PRP6-1	0.273	1.000	1.386
15	EP4, PRS2, ES2, EH2	PRP6-1	0.263	1.000	1.443
16	PRH3-1, PRS2, EP4, ES2, EH2	PRP6-1	0.258	1.000	1.444
17	EP4, POS1	PRP6-1	0.264	1.000	1.444
18	EV1-1, EV2-3	EH1	0.270	1.000	1.444
19	EP4, PRP8-2	PRP6-1	0.275	1.000	1.444
20	EP4, EH2, POS1	PRP6-1	0.281	1.000	1.256
21	PRH3-1, EP4, POS1	PRP6-1	0.287	1.000	1.256
22	PRH3-1, EP4, PRP8-2	PRP6-1	0.292	1.000	1.256
23	EP4, PRS2	PRH3-1	0.295	0.988	1.256
24	EP4, PRS2, ES2	PRH3-1	0.299	0.988	1.015
25	EP4, PRS2, ES2	PRH3-1	0.303	0.988	1.015
26	EP4, PRS2, EH2	PRH3-1	0.307	0.988	1.015
27	EP4, PRS2, ES2, PRP6-1	PRH3-1	0.309	0.985	1.015
28	EP4, PRS2, PRP6-1, EH2	PRH3-1	0.299	0.985	1.335
29	EP4, PRS2, ES2, EH2	PRH3-1	0.289	0.985	1.335
30	EP4, PRS2, ES2, PRP6-1, EH2	PRH3-1	0.280	0.985	1.335
31	PRP6-1, EV2-3	EH1	0.270	0.981	1.335
32	PRS2, PRP8-2, PRP6-1, EH2	PRH3-1	0.260	0.981	1.143
33	PRV1-3, EV2-3	EH1	0.262	0.975	1.293
34	PRP6-1, PRS2, EH2	PRH3-1	0.272	0.974	1.543
35	EV2-3, EP2	EH1	0.281	0.974	1.494
36	EV2-3, POH2	EH1	0.291	0.972	1.182
37	EP4, ES2	EH2	0.301	0.970	1.562

38	PRP6-1, EP4, ES2	EH2	0.311	0.968	1.562
39	PRP6-1, PRS2, ES2, EH2	PRH3-1	0.331	0.966	1.562
40	PRH3-1, EP4, ES2	EH2	0.318	0.964	1.544

3.4. Visualization of association rules from a holistic perspective

The 45 binomial association rules were visualized, and some of the association rules from the overall perspective were visualized as shown in Figure 2. The vertex code of the edge pointing represents the influencing factor term of modeling the relationship between Japanese yokai culture and folk beliefs, the edge represents the association rule of the relationship in the rule, the direction of the arrow represents the relationship between the leader and the successor in the association rule, and the tail of the arrow is the leader and points to the successor. The dot between the two factors indicates the support and confidence of the association rule, and the larger the size of the dot, the higher the confidence and the darker the color, the greater the support. The correlation between the influencing factors can be seen. The degree of modern commercial development (EV2-3) and the frequency of official literature (EH2) are independent of the whole factor relationship diagram, and the correlation with other influencing factors is not strong, indicating that in the modeling of the relationship between Japanese yokai culture and folk beliefs, commercial development is a phenomenon of modern civilization, and there is a temporal and spatial separation between the mining target of the association rules of historical folk beliefs, and the frequency of official literature is difficult to effectively record the integration of people's participation and beliefs, indicating that the rules are unreliable.

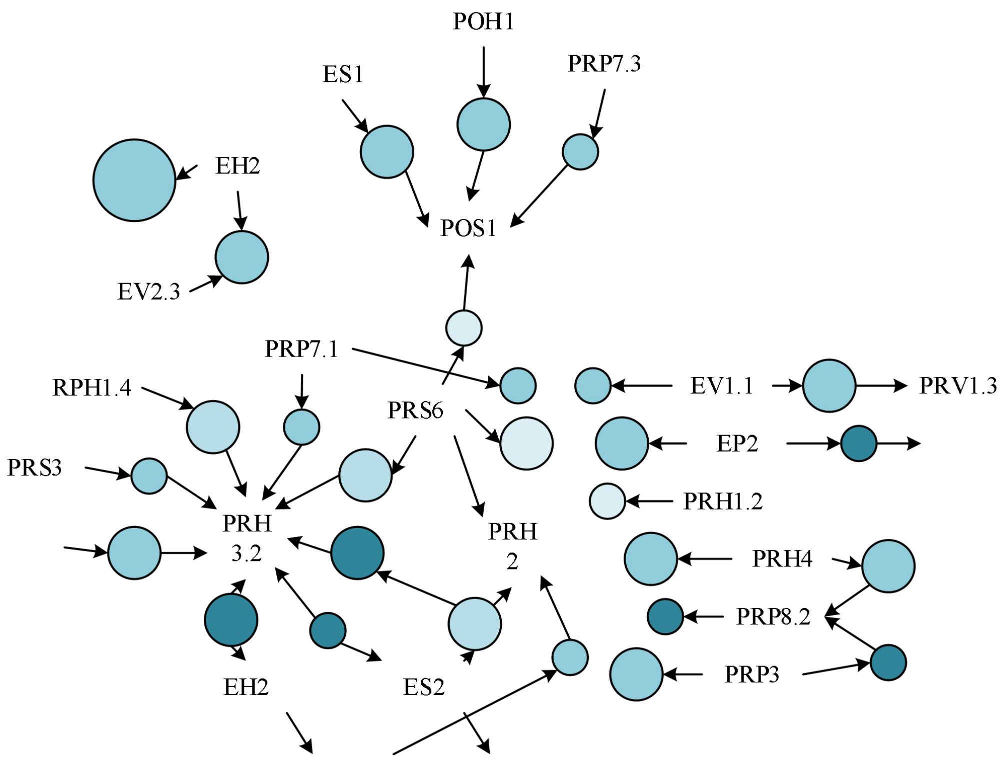

Figure 2. Partial correlation rules visualization.

4. Network Construction and Analysis of Japanese Yokai Culture and Folk Beliefs

4.1. Japanese Yokai Culture and the Construction of Folk Belief Networks

Calculate the co-occurrence frequency of 40 Japanese monster culture and folk belief characteristics in different corpora. If a set of keywords appears multiple times in a survey report, it is counted as one co-occurrence. The final co-occurrence matrix of Japanese monster culture and folk belief characteristics is shown in Table 3. The larger the co-occurrence value between two characteristics, the stronger their association. Based on the constructed co-occurrence matrix of Japanese monster culture and folk belief characteristics. The co-occurrence value between Feature 1 and Feature 2 reached 95, indicating a strong association between these two features.

In the network diagram of Japanese yokai culture and folk belief relationships, nodes represent features of Japanese yokai culture and folk beliefs, and the lines connecting the nodes indicate the relationships between the features. Larger nodes indicate that the feature appears more frequently in the entire corpus and has a greater influence on the relationships between Japanese yokai culture and folk beliefs. Thicker and denser lines between nodes indicate a closer connection between the features.

Table 3. Partial eigencoactive matrix.

N	1	2	3	4	5	6	7	8	9	10
1	0	95	89	65	75	62	65	68	45	48
2	95	0	82	60	65	64	53	40	38	39
3	89	82	0	61	58	59	57	56	55	62
4	65	60	61	0	45	43	46	47	42	45
5	75	65	58	45	0	40	38	39	32	30
6	62	64	59	43	40	0	35	34	33	31
7	65	53	57	46	38	35	0	29	28	26
8	68	40	56	47	39	34	29	0	28	25
9	45	38	55	42	32	33	28	28	0	21
10	48	39	62	45	30	31	26	25	21	0

4.2. Small World Characteristics Analysis

Using average shortest path length and clustering coefficient to quantify the small-world characteristics of the network of relationships between Japanese yokai culture and folk beliefs. The network of relationships between Japanese yokai culture and folk beliefs constructed by the author has an average path length defined as the average of the shortest distances between any two cultural influence factors. The average path length of a network reflects its propagation speed; a larger average path length indicates slower interaction between cultural elements, suggesting weaker connections between influence factors. Conversely, a smaller average path length indicates easier multi-factor cultural coupling. Using Pajek software, the average shortest path length of the network was calculated to be 1.05. The maximum distance between any two nodes in the network is called the network's diameter. Using Pajek software, the maximum distance between two nodes in the network was calculated to be 2.1, obtained from the node “monster attributes” to the node “path of introduction from the South Seas.” The clustering coefficient of the network is used to quantify the clustering characteristics of the nodes in the network. The calculated clustering coefficient values for all nodes are greater than 0.9, far exceeding the clustering coefficient values of random networks of the same scale. This indicates that, on average, every two nodes in the network are connected. The network has a small average shortest path length and a high clustering coefficient, indicating that the network of relationships between Japanese demon culture and folk beliefs exhibits small-world characteristics, forming a tightly interwoven network of belief dissemination.

4.3. Network Centrality Analysis

An analysis of the centrality of nodes in the network of relationships between Japanese yokai culture and folk beliefs can be used to measure the importance of keywords within the network. The primary metrics for calculating network centrality include degree centrality, betweenness centrality, and closeness centrality. The centrality values for each feature are shown in Table 4. Nodes 1, 2, 3, 8, and 10 have the highest closeness centrality z values, all at 1.000, and their degree centrality values are also higher than those of other features. The higher the degree centrality of a node, the more important it is in the entire network. Considering the frequency of co-occurrence of each feature in different Japanese folklore and folk belief-related survey documents, node degree values are assigned weights. According to the degree centrality analysis, the most important factors influencing the relationship between Japanese folklore and folk beliefs are the types of folklore, regional distribution characteristics, historical periods, cultural transmission media, and the degree of belief integration, as indicated by nodes 1, 2, 3, 8, and 10. Closeness centrality reflects the proximity of a node to other nodes. A smaller value indicates a shorter path to other nodes, suggesting a closer connection between this node and others. Based on closeness centrality and intermediary centrality analysis, it is evident that factors such as yokai types, regional distribution characteristics, historical periods, cultural transmission media, and degree of belief integration significantly influence the network, with close relationships between features. This reflects the diverse influence pathways and short causal chains characteristic of the relationship between Japanese yokai culture and folk beliefs.

Table 4. Central nature of characteristics.

N	Centrality	Intermediate center	Proximity center	Node number	Centrality	Intermediate center	Proximity center
1	751	0.004	1.000	20	541	0.003	0.945
2	688	0.003	1.000	21	411	0.003	0.945
3	693	0.004	1.000	22	182	0.001	0.922
4	549	0.004	0.971	23	451	0.002	0.946
5	245	0.004	0.971	24	192	0.004	0.702
6	512	0.001	0.924	25	429	0.002	0.941
7	595	0.004	0.973	26	105	0.002	0.751
8	854	0.004	1.000	27	342	0.004	0.971
9	312	0.002	0.908	28	62	0.001	0.751
10	841	0.004	1.000	29	541	0.003	0.942
11	226	0.001	0.901	30	606	0.002	0.945
12	81	0.001	0.842	31	64	0.000	0.698
13	92	0.000	0.821	32	414	0.003	0.974
14	533	0.003	0.974	33	122	0.002	0.687
15	675	0.004	0.974	34	256	0.001	0.881
16	378	0.004	0.912	35	118	0.001	0.881
17	285	0.004	0.879	36	191	0.001	0.841
18	89	0.002	0.902	37	584	0.003	0.949
19	189	0.002	0.654	38	115	0.001	0.714
20	510	0.003	0.974	39	606	0.002	0.945

4.4. Cohesion Characteristics Analysis

The cohesiveness of nodes reflects the interconnectivity of nodes within a network. Within a cohesive subgroup, nodes are connected by relatively strong and direct ties. The more connections between nodes, the more stable the structure and the stronger the cohesiveness. The k-core decomposition method is employed to identify cohesive subgroups that characterize the relationship between Japanese yokai culture and folk beliefs, thereby analyzing the cohesiveness among these features.

Conducting a cohesive characteristic analysis of the network of features related to Japanese yokai culture and folk beliefs allows for simultaneous analysis of core causal factors and their closely related peripheral causal factors, enabling targeted control of the mutual influence between core and peripheral factors.

The k-core decomposition process is conducted in an outward-to-inward expansion manner. The smallest core nodes are typically located at the outermost layer of the network, while the largest core nodes are located at the innermost layer. The frequency distribution of k-core nodes is shown in Table 5, which is divided into 10 subcategories. The highest core in the k-core structure network is the 40-core, and the 40-core structure constitutes the largest connected subgraph of the network, which represents the core group of the network. The core characteristics of the Japanese monster culture and folk belief relationship network are: monster type (1), regional distribution characteristics (2), historical period (3), cultural transmission medium (8), degree of belief integration (10), monster good-evil attributes (16), and natural element dependency attributes (28). These characteristics exhibit strong clustering, with each characteristic directly linked to others. Therefore, the primary influencing factors of the network of relationships between Japanese demon culture and folk beliefs are typically the result of the interplay of multiple factors such as demon types, regional distribution characteristics, historical periods, and cultural transmission vehicles.

Table 5. K-nuclear node frequency distribution.

K	Number of nodes	Node number
23	2	4, 26
25	1	25
28	4	5, 15, 24, 35
30	2	9, 14
33	8	13, 17, 32, 34, 36, 37, 38, 39
35	3	6, 7, 23
36	2	21, 22
38	6	11, 12, 18, 30, 31, 33
39	5	19, 20, 27, 29, 40
40	7	1, 2, 3, 8, 10, 16, 28

5. Conclusion

This paper uses association rule mining and social network analysis to build a model of the relationship between Japanese yokai culture and folk beliefs, revealing the underlying logic between the two.

Based on 370 association rules, three rules with a support value of 0.5 or higher were selected, confirming that there is a frequent association between yokai types, cultural transmission carriers, and regional distribution characteristics. In high-confidence association rule mining, only four influencing factors were identified as subsequent items, and the rule confidence levels of all four factors were above 0.96. This indicates that under the premise of the leading item's appearance, the degree of Buddhist faith integration, regional distribution characteristics, yokai type: kappa, and the frequency of official document records have a higher probability of appearing in studies on the relationship between Japanese yokai culture and folk beliefs.

The network constructed in this paper exhibits small-world characteristics, with an average shortest path of only 1.05, ensuring efficient interaction among the influencing factors, and a maximum path of 2.1, revealing the rapid integration of Japanese yokai cultural attributes and South Asian cultural influences introduced into the local culture.

The cohesive characteristics of the network of Japanese yokai culture and folk belief reveal that the core features of the network are: yokai types, regional distribution characteristics, historical periods, cultural transmission carriers, degree of faith integration, yokai good-evil attributes, and natural element dependency attributes. The revelation of these core features makes Japanese yokai culture and folk belief a scientifically predictable research system, providing a reference method for the protection of East Asian cultural heritage.

REFERENCES

[1] Oguma, M. (2016). The Study of Japan through Japanese Folklore Studies Anthropology and Folklore in Japan. Japanese review of cultural anthropology, 16, 237-250.

[2] Xu, Y., & Hamamura, T. (2014). Folk beliefs of cultural changes in China. Frontiers in psychology, 5, 1066.

[3] Korman, D. Z. (2009). Eliminativism and the challenge from folk belief. Noûs, 43(2), 242-264.

[4] Papp, Z. (2010). Traditional monster imagery in manga, anime and Japanese cinema. Global Oriental.
[5] Mittman, A. S., & Dendle, P. J. (2017). Early Modern Past to Postmodern Future: Changing Discourses of Japanese Monsters. In The Ashgate Research Companion to Monsters and the Monstrous (pp. 173-190). Routledge.
[6] Lundby, K. (2016). Mediatizing Faith: Digital Storytelling on the Unspoken 1. In Mediating Faiths (pp. 49-62). Routledge.
[7] Hong, N. (2022). Digital‐Media‐Based Interaction and Dissemination of Traditional Culture Integrating Using Social Media Data Analytics. Computational Intelligence and Neuroscience, 2022(1), 5846451.
[8] Alshawi, A. S., Alghaithi, S. A., Alrayssi, J. A., Alsuwaidi, A. E. E., & Hamzah, A. R. B. H. (2022). The Impact of the Use of Digital Media in the Radio on the Dissemination of Islamic Culture. International Journal of Health Sciences, (III), 651-658.
[9] Dahlan, N. F. (2024). Media and Culture: The role of media and social media in the formation and dissemination of popular, local, global and international culture. Journal of Studies on Art, Culture and Society, 13-17.
[10] Machmud, M., Rijal, N. K., & Anugrah, M. (2024). Traditional and Digital Communication Media Publication Trends: A Systematic Literature Review. Journal of Ecohumanism, 3(8), 8143-8154.
[11] Yan, Z., Lim, C. K., Hu, D., Ahmed, M. F., Halim, S. A., & Li, L. (2025). Construction of digital dissemination effects evaluation indicator system of traditional techniques of intangible cultural heritage. npj Heritage Science, 13(1), 1-11.
[12] Jin, P., & Liu, Y. (2022). Fluid space: Digitisation of cultural heritage and its media dissemination. Telematics and Informatics Reports, 8, 100022.
[13] Yang, R., & Abidin, S. Z. (2022, May). Research on the dissemination and application of computer 3D technology in the process of intangible cultural heritage display. In 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI) (pp. 1339-1342). IEEE.
[14] Liu, Z. (2025). The construction of a digital dissemination platform for the intangible cultural heritage using convolutional neural network models. Heliyon, 11(1).
[15] Ryotaro Koga,Tomohiro Tsuchida,Chihiro Tsunoda,Hikaru Kataoka,Shoya Suenaga & Satoru Goto. (2025). Physicochemical properties of diclofenac regulated by

its intermolecular interactions with caffeine evaluated via vector space model, nuclear magnetic resonance, and Fourier-transform infrared analyses. Materials Chemistry and Physics,333,130158-130158.
[16] Abubakar Salisu Bashir,Abdulkadir Abubakar Bichi & Alhassan Adamu. (2024). Automatic construction of generic Hausa language stop words list using term frequency-inverse document frequency. Journal of Electrical Systems and Information Technology,11(1),58-58.
[17] Shujie Wang. (2025). Social network analysis enhances job recommender system - PageRank algorithm in the field of employment. Journal of Computational Methods in Sciences and Engineering,25(4),3210-3223.
[18] original/au/au-aff. (2018). Evolution of condensed subgroup of political participation based on mobile phone. Multimedia Tools and Applications,77(4),4271-4282.
[19] Jing Zhang. (2025). Data-driven development of entrepreneurial ecosystem based on the Apriori algorithm. Journal of Computational Methods in Sciences and Engineering,25(4),3224-3238.