

Retraction

Retracted: Research on the Popularity Prediction of Multimedia Network Information Based on Fast K Neighbor Algorithm

Journal of Electrical and Computer Engineering

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Journal of Electrical and Computer Engineering. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.


The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] X. Mao and F. Xu, "Research on the Popularity Prediction of Multimedia Network Information Based on Fast K Neighbor Algorithm," *Journal of Electrical and Computer Engineering*, vol. 2022, Article ID 8745492, 11 pages, 2022.

Research Article

Research on the Popularity Prediction of Multimedia Network Information Based on Fast K Neighbor Algorithm

Xiaodong Mao¹ and Fei Xu² 

¹Yongcheng Vocational College, Yongcheng 476600, China

²Beijing Taohuadao Information Technology Co.,Ltd., Beijing, China

Correspondence should be addressed to Fei Xu; xufeifei@mail.ustc.edu.cn

Received 25 February 2022; Revised 15 March 2022; Accepted 1 April 2022; Published 28 April 2022

Academic Editor: Wei Liu

Copyright © 2022 Xiaodong Mao and Fei Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

From the perspective of practical application, information popularity prediction is of positive significance for corporate marketing, advertising, traffic control, and risk management. This paper combines the fast K-nearest neighbor algorithm to predict and analyze the popularity of multimedia network information and improves the nonindependent and identically distributed KNN classification algorithm. Moreover, this paper proves that it is a superior measurement method when considering the nonindependent and identical distribution among data objects to measure similarity and the improved CS_KNN algorithm can greatly improve the classification performance. Finally, this paper constructs a prediction model of multimedia network information popularity based on the fast K neighbor algorithm. Through the experimental research results, it can be seen that the prediction effect of the multimedia network information popularity prediction system based on the fast K neighbor algorithm proposed in this study is very good.

1. Introduction

Traditional media has developed rapidly under the drive of global information networks and computer technology, and many new tools have begun to be put into application, which has greatly promoted the development of the media industry. Related studies have also found that the time people spend on general reading and newspaper reading is decreasing, and their attention to paper media is decreasing, while the time spent on online media is increasing [1]. These phenomena also show that new media has occupied an important position in the media industry, and people's methods of obtaining media information and learning habits have undergone important changes. In order to meet the news needs of the increasingly large-scale netizens, the online news topics of major websites in our country have been made more and more specific and detailed, which has formed a boom in online news topics [2]. There are many ways to spread online media, and the corresponding display methods are also rich and diverse. For example, it can be displayed in traditional text, pictures, and can also be displayed through animation videos. With the

current rapid development of network technology, new media models will continue to emerge and be applied to the field of news media. The large number of applications of these network communication technologies will also have a great impact on the mode of network news dissemination, and its dissemination effect has been greatly improved [3].

At present, it is not difficult to find that even the reading volume of articles with similar content will be very different. If various online media want to improve the quality of news and increase the retention of users, they must strengthen the supervision of the news itself at the source. Reading volume of online news is one of the indicators to measure the quality of news. As far as the network news itself is concerned, we can monitor the characteristic information such as the number of images, the number of videos, the number of keywords, and the length of the news in the body of a piece of network news. Through the characteristic information of these online news, we can analyze its popularity, and through the analysis results, we can make suggestions for improvement on the way of online media editors' news writing and the priority of user recommendations.

Based on the above analysis, this paper combines the fast K-proximity algorithm to predict and analyze the popularity of multimedia networks, constructs a corresponding intelligent model, and verifies the performance of the model to improve the effect of subsequent multimedia network information popularity prediction.

2. Related Work

Information popularity prediction in social networks is an important branch of information dissemination prediction, which has important research and application value [4]. Prediction of information popularity for influencing factors before release mainly considers the inherent dissemination influence of the information itself, predicts the popularity of the information before it is released, and provides support for controlling the reasonable release of information. Early factors considered include publisher influence, text content, etc. Literature [5] analyzes and models the inherent multidimensional attributes of news. Starting from multiple influencing factors such as news categories, article language characteristics, and whether users are authenticated, it uses regression algorithms to combine the early multidimensional attributes of news, and uses classification algorithms to predict future issues. The popularity of the content in this article has a prediction accuracy of 84%. Literature [6] deeply studied the intrinsic properties of blog content, that is, the impact of content eloquence and novelty on information popularity, using the number of links to measure novelty, and the length of content to measure eloquence. The research in [7] shows that the fewer blog citation links and the longer the content, the more likely it is to gain higher popularity in the later period. Literature [8] studies the influence of some early intuitive attributes on the popularity of news, mainly including influence and text content. Literature [9] uses a classifier to predict whether news will be commented, and then uses another classifier to predict the number of comments for those news that are predicted to be commented. The analysis and forecasting of influencing factors after the release mainly focuses on the establishment of prediction models for the attributes of participating users and time series factors after the information is released. Literature [10] found through experiments that most of the follow-up popularity of information on Digg posts comes from the fans of people who liked before. Literature [11] found that Weibo information has a similar phenomenon. It extracts the characteristics of the forwarding user and the interaction characteristics between the forwarding user and the fans. Through time slicing, a classification model is used to predict whether the fans of the forwarding user will participate in the next moment. Come to the topic, and then perceive the situation of information dissemination. Literature [12] mines the deep social influence of information through comment content, and mines the time factor of information popularity through comment time, and proposes a ranking model based on bipartite graph and regularization to predict the popularity of information in the future. Literature [12] uses the number of reposts to measure the popularity of microblog information, extracts

multidimensional reposting factors such as user reposting interest and user activity, and takes hot information in the current context as a consideration factor that affects information popularity, and uses the classification model to solve the problem of whether users forward or not. Literature [13] combines game theory to extract external and internal attributes to quantify the infection rate of SIR, uses SIR to construct a state model of information dissemination, and uses state I to perceive the trend of the popularity of information dissemination. In addition, some people comprehensively consider the influencing factors before and after the release, convert the popularity prediction into the prediction of the number of reposts, and train the classification model to realize the popularity prediction by extracting attributes. According to the three attributes of user forwarding, the emotion of the information content, and the user's interest, Liu Hehe and others constructed a dynamic characteristic user forwarding behavior prediction model to predict the number of topic information forwarded at a time [14].

In the research of statistical information popularity, statistical methods are mainly used to explore some regularities of information popularity changes and to achieve information popularity prediction. Literature [15] studies the growth and decay trend of online information popularity over time, uses a K-Spectral Centroid (K-SC) algorithm to cluster the popularity of time similarity, and divides the trend of popularity into six categories. Literature [16] found that the trend of popularity is related to the choice of time span. In addition to the above six popularity evolution models, there are complex repetitions and fluctuations in the popularity trend. Literature [17] analyzed the information popularity growth pattern of YouTube videos and Digg stories and found that there is a strong logarithmic relationship between early and early popularity and future popularity. Literature [18] considers the amount of forwarding caused by external indirect attention users, based on the traditional Susceptible-Infected-Susceptible (SIS) model, newly added the external access state E (External), constructs the infectious disease SISE propagation model, and extracts each time sequence. According to the value of the corresponding state, train the model to find the infection rate, realize the prediction of the number of forwarding, and sense the change of the information popularity situation. Literature [19] adds state E based on the traditional SIS model. Literature [20] optimized the initial value and threshold of backpropagation (BP) neural network through genetic algorithm, and constructed a network public opinion crisis early warning model based on BP neural network and genetic algorithm.

3. Prediction Algorithm Based on Fast K Nearest Neighbor Algorithm

The "majority voting" method is a simple discriminant rule, and its basic idea is to return the class label with the most categories among the k neighbors as the prediction result. We set x as the sample to be tested and x_i as the k nearest neighbors of x . The total number of training sample

categories is z , and the category set is $C = \{C_1, \dots, C_t, \dots, C_z\}$, where $t \in [1, z]$. Then the function to distinguish the category of the sample to be tested x is

$$W(x, c_t) = \sum_{i=1}^k \varphi(x_i, c_t). \quad (1)$$

Here, there are

$$\varphi(x_i, c_t) = \begin{cases} 1, & x_i \in c_t, \\ 0, & x_i \notin c_t. \end{cases} \quad (2)$$

That is, if $x_i \in c_t$, $\varphi(x_i, c_t) = 1$; otherwise, $\varphi(x_i, c_t) = 0$. The decision rules are the following:

$$\hat{c}(x) = \arg \max(W(x, c_t)). \quad (3)$$

Here, the return value of the algorithm is $\hat{c}(x)$'s estimate of $c(x)$, which represents the prediction category of the test instance x . Although the majority voting rule is simple in thinking and easy to implement, it only intuitively considers the number of each category in the sample for statistics, and does not consider other implicit information in the k nearest neighbors. Important information such as the distance and sample distribution implied in the neighbor samples will have a great impact on the classification effect.

The basic idea of the distance-weighted voting decision-making method is when making a decision, the quality of neighbors is expressed by the size of the assigned weight. Compared with traditional voting decision-making methods, this method uses the difference in similarity between neighbors and the sample to be tested, and it uses known information to a certain extent.

The category discriminant function is as follows:

$$W(x, c_t) = \sum_{i=1}^k w_i \varphi(x_i, c_t). \quad (4)$$

Here, x_i is the k nearest neighbors of the sample x to be tested, and $\varphi(x_i, c_t)$ is defined as in formula (2). The weight function is generally a mapping function of distance, which is defined here as follows:

$$Vsim(Y, X_i) = \left(\frac{\text{sim}(Y, X_i) - \min_{X_f \text{ is } K \text{ neighborsof } Y} \text{sim}(Y, X_f)}{\max_{X_f \text{ is } K \text{ neighborsof } Y} \text{sim}(Y, X_f) - \min_{X_f \text{ is } K \text{ neighborsof } Y} \text{sim}(Y, X_f)} \right)^{2/\tau-1}. \quad (7)$$

Here, X_f is the k nearest neighbors of Y , and τ is the parameter, which is 2 here.

The similarity function between the two samples X_a and X_b in the data set is as follows:

$$\text{sim}(X_a, X_b) = \left(\sqrt{\sum_{j=1}^M (X_{aj} - X_{bj})^2} \right)^{-1}. \quad (8)$$

$$w_i = \frac{1}{d(x_i, x)^2}. \quad (5)$$

The decision rules are the following:

$$\hat{c}(x) = \arg \max(W(x, c_t)). \quad (6)$$

That is, the predicted category of the sample x to be tested is $\hat{c}(x)$.

From the above analysis, it can be seen that after the traditional KNN algorithm selects k neighbors, it is assumed that the k neighbors have the same influence on the sample to be tested as in the decision-making process. The problem with this method is that it ignores the influence of the similarity between different neighbors and the sample to be tested on the classification, which greatly affects the accuracy of the KNN classifier. Figure 1 shows an example of two classifications. In the figure, the circle represents category A, and the rounded rectangle represents category B. Observation shows that the point X to be tested should be of type A, but if the traditional decision-making method is used, the predicted classification result will be of type B. Therefore, ignoring the quality of neighbors in traditional decision-making methods will lead to the misjudgment of the classifier to a certain extent.

In the decision rules, different weights are assigned to different neighbors according to the distance or similarity function between the test sample and the neighbor samples. This idea was first proposed by Dudani. Accordingly, this algorithm introduces the concept of nearest neighbor support in decision-making. It uses the similarity value of the neighboring point X and the measured point Y in the decision domain to construct the nearest neighbor support of X . If the similarity value between X and Y is greater, the neighbor support is greater. This shows that the neighbor X has a higher degree of influence on the decision result of Y .

Definition 1. The calculation formula for the neighbor support $Vsim(Y, X_i)$ of the neighbor X_i of the sample Y to be tested is as follows:

Here, X_{aj} and X_{bj} represent the characteristic value of the characteristic j of the sample X_a and X_b . It can be seen from the formula that the higher the similarity value of the neighboring points X and Y , the higher the $Vsim(Y, X_i)$.

According to the KNN decision rules, it can be found that the classifier ignores the distribution of samples in the space when making decisions. Ignoring the sample distribution has the following two problems: one is that when the

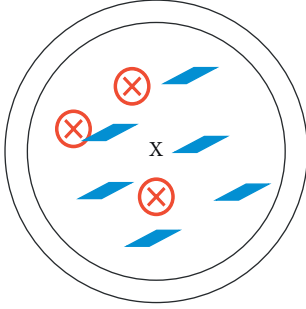


FIGURE 1: Two-classification example.

number of sample categories is large, the classifier will tend to make a misclassification of large samples; the second is when the local density of the adjacent area of the data to be measured is uneven, the data to be measured easily classified as a high-density category. In order to improve the problem that traditional algorithms are affected by sample distribution, this paper introduces class credibility in the decision-making stage. The basic idea is to add intraclass local factors in decision-making, and then improve the classification accuracy by considering the distribution of samples.

When the sample distribution density is not uniform, the classification effect of the traditional KNN algorithm will decrease to a certain extent. Figure 2 is an example of two classifications, in which the rounded rectangle represents category A, the circle represents category B, and the distribution density of the two categories is uneven. It can be seen from the figure that among the k neighbors of the point X to be measured, the denser type A always accounts for the majority of neighbors. Therefore, according to the traditional KNN classifier, no matter what the value of k is, the sample to be tested tends to be classified into the dense A class. Therefore, it is possible to improve classification accuracy by considering sample distribution information when making decisions. In this paper, the intraclass local factor is set by considering the distribution of the sample to be tested and the neighbors in the various e -neighborhoods.

Definition 2. The intraclass local factor $\rho(c_t)$ in the intraclass sample neighborhood e of class c is

$$\rho(c_t) = \frac{1}{\varepsilon} \sum_{b=1, b \neq a}^{\varepsilon} \sum_{a=1}^{\varepsilon} \text{sim}(X_a, X_b). \quad (9)$$

Here, X_a and X_b are samples in the sample neighborhood e of the sample Y belonging to c , in the sample to be tested. Among them, the sample neighborhood z within the class belonging to c , refers to the neighborhood range of z neighbors of the test point Y belonging to the class c_t .

Definition 3. The class credibility of class c_t is

$$\mu(c_t) = e^{-\rho(c_t)}. \quad (10)$$

Here, $\rho(c_t)$ is the intraclass local factor of class c_t .

The improved ND_KNN algorithm steps are divided into two stages. The algorithm flow chart is shown in

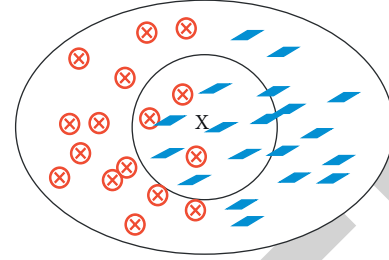


FIGURE 2: Example of uneven density.

Figure 3. The specific steps are as follows: the input is sample $Y = \{Y_1, Y_2, \dots, Y_N\}$, to be tested, where $l \in (1, N)$. This study sets up an M -dimensional training sample set $X = \{X_1, X_2, \dots, X_n\}$, the total number of samples is n , and $h \in (1, n)$. The training sample category set is $C = \{C_1, C_2, \dots, C_n\}$, and the total number of categories is z .

The output is the predicted category C of the sample Y to be tested.

The first stage is that the algorithm calculates the k nearest neighbors of the test sample:

Step 1: it standardizes the characteristics of the original data set and stores it.

Step 2: it sets the initial value of parameter k , and the final value of this experiment is selected according to the experimental results.

Step 3: it chooses the reciprocal Euclidean distance as the similarity measure function $\text{sim}(-)$. It traverses the training set and calculates the similarity between Y and the current instance.

Step 4: the traversal ends, and the most similar k tuples are stored in the set $T = \{T_1, T_2, \dots, T_n\}$, $f \in (1, k)$ according to the similarity.

The second stage is for the selected k neighbors, and it determines the category of Y according to the category judgment function.

Step 5: it calculates the category judgment function $CW(Y, c_t)$ of the sample Y to be tested. Then, the category determination function for the sample Y to be tested belongs to category c_t is

$$CW(Y, c_t) = \mu(c_t) \sum_{i=1}^k V \text{sim}(Y, X_i) \varphi(X_i, c_t). \quad (11)$$

Here, $\varphi(X_i, c_t)$ is the same formula. It can be seen from the formula that the larger the category determination function, the greater the probability that Y belongs to the category c_t .

Step 6: it classifies the sample Y to be tested as the category with the largest category determination function, and the decision formula is

$$c(Y) = \text{argmax} CW(Y, c_t). \quad (12)$$

Step 7: it repeats the above steps until all samples in the sample set to be tested are classified.



FIGURE 3: ND_KNN flow chart.

In traditional research, distance is often used to determine the similarity between objects. On the one hand, distance function and similarity function, as two common concepts with opposite meanings, are two closely related universal measures. On the other hand, there is a certain internal connection between the similarity measurement function and the distance function, which can be derived from each other.

We assume that there is a certain mapping relationship D between the two and define it as follows:

$$D: d(x, y) \longrightarrow \text{sim}(x, y). \quad (13)$$

Here, $d(x,y)$ represents the distance formula between x and y , and $\text{sim}(x, y)$ represents the similarity function between x and y . Then the mapping relationship D between the two needs to meet the following conditions:

$$\text{If } d(x_1, y_1) > d(x_2, y_2), \text{ then } \text{sim}(x_1, y_1) \leq \text{sim}(x_2, y_2)$$

$$\text{If } d(x_1, y_1) < d(x_2, y_2), \text{ then } \text{sim}(x_1, y_1) \geq \text{sim}(x_2, y_2)$$

$$\text{If } d(x_1, y_1) = d(x_2, y_2), \text{ then } \text{sim}(x_1, y_1) = \text{sim}(x_2, y_2)$$

Here, D represents the transfer relationship between the similarity function and the distance function.

We set up the space to be denoted as Ω , x , y , and z are any three points in Ω . The distance function $d(x, y)$, as a function of measuring the distance of an object, needs to satisfy the following properties at the same time:

- (1) Nonnegativity is if $d(x, y) \geq 0$, then $\forall x, y \in \Omega$. That is, the distance between any two objects cannot be negative.
- (2) The symmetry is that if $d(x, y) = d(y, x)$, then $\forall x, y \in \Omega$. That is, the distance between x and y is the same as the value between y and x .
- (3) The identity is that if $d(x, x) = 0$, then $\forall x \in \Omega$. That is, the same sample is completely similar.

- (4) The triangle inequality is that if $d(x, y) \leq d(x, z) + d(z, y)$, then $\forall x, y, z \in \Omega$. That is, the distance between the object x and y is not greater than the sum of the distances between x, z , and y, z .

For any clustering or classification problem in data mining, determining the appropriate measurement method according to the characteristics of the data set, such as the data type, sample distribution, and algorithm process, is the core issue. The following are three commonly used distance formulas.

3.1. Euclidean Distance. We set x and y to be two points in the n -dimensional feature space, then the Euclidean distance between x and y is

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (14)$$

We set x and y to be two points in the n -dimensional feature space, then the cosine similarity between x and y is

$$\cos \theta = \frac{x \cdot y}{|x| \cdot |y|} = \frac{\sum_{k=1}^n x_k \cdot y_k}{\sqrt{\sum_{k=1}^n x_k^2} \sqrt{\sum_{k=1}^n y_k^2}}. \quad (15)$$

The cosine similarity is mapped to the n -dimensional vector space by mapping the sample object, and the similarity is expressed by the cosine value of the angle between the cosine values of the two vectors. Usually, the cosine distance range is $[-1, 1]$. The closer the angle value is to 0, the closer the two vectors are. In extreme cases, if the cosine value is 1, it means that the two objects are completely similar. The cosine similarity is not sensitive to the real number value, and it pays attention to the difference in direction. Compared with Euclidean distance, it is more suitable for text-type object similarity, and then completes the classification of text-type objects.

3.2. Mahalanobis Distance. We set up the data set $X = \{X_1, X_2, \dots, X_m\}$, S is the covariance matrix of the data set, and \bar{X} is the mean value. Then the Mahalanobis distance from X to u is

$$D(X) = \sqrt{(X - \mu)^T S^{-1} (X - \mu)}. \quad (16)$$

3.3. Feature Normalization. When the value range of the feature value is quite different, such as "height," "age," "income," and these features have different value ranges. If not processed, the output result will be largely "income." It is influenced by this characteristic. The common way is to scale the data proportionally before the similarity measurement so that the feature value is within a range of value. There are generally two ways to standardize: one is normalization with their respective mean and variance estimates, and the other is minimum and maximum normalization. The formula

introduces the minimum and maximum normalization method, and here is the average method. For the j -th dimension feature containing n data, there are

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_k}. \quad (17)$$

We have the following:

$$\sigma_k^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}. \quad (18)$$

3.4. Feature Weighting. One of the classic methods to solve the curse of dimensionality is to weight features when measuring similarity. We assume that in the feature space, the idea of feature weighting is to scale the coordinate axis representing a feature according to the degree of relevance, including lengthening the coordinate axis of strong correlation features and shortening the coordinate axis of weak correlation features, and the common feature weighting method is

$$D(Y, X^i) = \sqrt{\sum_{i=1}^n (y_i - x_{ij})^2}. \quad (19)$$

The following introduces a feature weighting method that determines the weight coefficient by measuring the information gain. In order to enhance the strength of the feature's impact on classification and optimize the distance measurement, this method introduces information gain and information to "enlarge" key features, "shrink" secondary features, and "remove" irrelevant features. The steps for calculating feature weight coefficients are as follows:

First, it calculates the information gain value $\text{Gain}(i)$ of each feature in the n -dimensional space, $0 \leq i \leq n$.

$$\text{Gain}(A) = E(S) - E(S|A) \geq 0. \quad (20)$$

Here, the characteristic value of attribute A can be taken as a_1, a_2, \dots, a_m . These values can divide the sample set S into m subsets $\{S_1, S_2, \dots, S_m\}$:

$$E(A) = \sum_{i=2}^n p(a_i) \log_b \frac{1}{p(a_i)}. \quad (21)$$

Here, b is taken as 2, and $p(a_i)$ is the probability of occurrence of event a_i . The conditions for classifying by attribute A are

$$E(S|A) = \sum_{j=1}^m p(S_j) \sum_{i=1}^n p(C_i|S_j) \log_b \frac{1}{p(C_i|S_j)}. \quad (22)$$

Then, it calculates the feature weight coefficient w_i , the formula is as follows:

$$w_i = \frac{\text{Gain}(i)}{\sum_{i=1}^n \text{Gain}(i)} \cdot e^{\text{Gain}(i)}. \quad (23)$$

Finally, the feature weight coefficient w_i is used to measure the distance. It can be seen from the formula that the

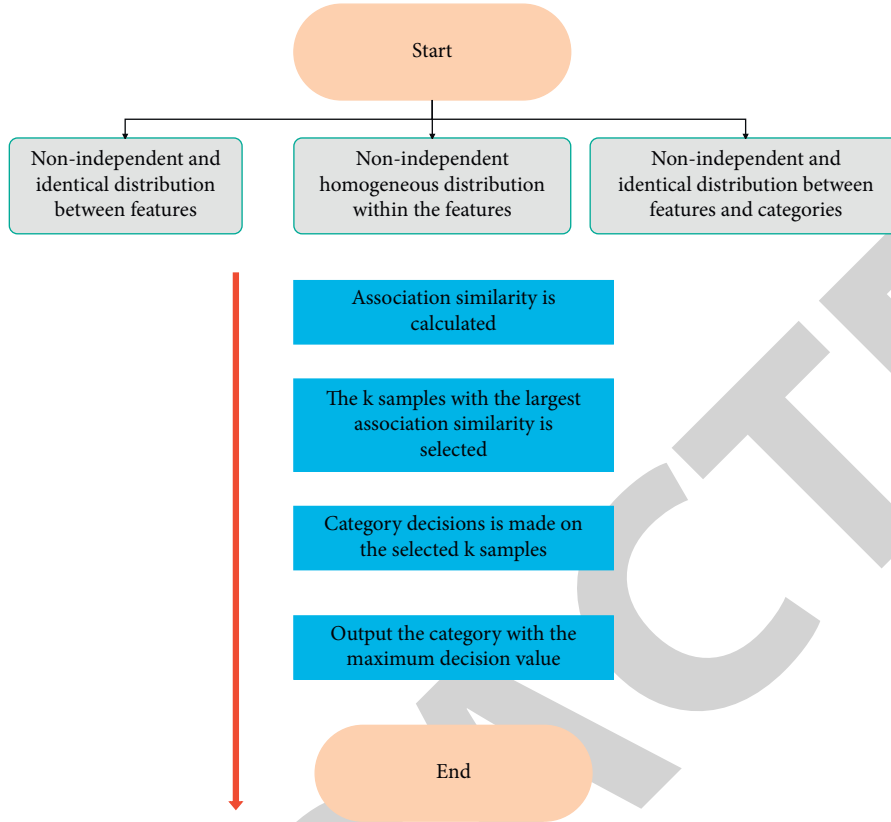


FIGURE 4: CS_KNN algorithm flow chart.

feature w_i with a large $\text{Gain}(i)$ is also large, on the contrary, the w_i value is small with a small $\text{Gain}(i)$. The time complexity of the improved KNN algorithm does not increase.

We assume that a certain sample Y in the training set belongs to category A , then the stronger a certain feature of Y contributes to the classification, the greater its weight value. In addition, the class attribute characteristic value of the characteristics belonging to the B sample is different from the class attribute characteristic value of the A sample. The idea of nonindependent and identical distribution between features and categories in this study is to map the distribution of different features on similar samples to the degree of classification influence, and by considering the classification strength of different features for the same category, different class weight coefficients are formed. The class feature weight coefficient α_{tj} between features and categories is shown in Definition 1.

Definition 4. The class feature weight of the sample X belonging to the t class on the feature is

$$\alpha_{tj} = \left(\sum_{d=1}^z \frac{\sqrt{\sum_{X_i \in t} (X_{ij} - Y_{tj})^2}}{\sqrt{\sum_{X_e \in d} (X_{ej} - Y_{dj})^2}} \right)^{-1}. \quad (24)$$

Amid them, there are

$$Y_{tj} = \frac{1}{\text{Num}(t)} \sum_{X_i \in t} X_{ij}. \quad (25)$$

$\text{Num}(t)$ represents the number of samples belonging in the training set, and the total number of classes is z . $X_e \in d$ and $X_i \in t$, respectively, indicate that X_e and X_i are training samples belonging to class d and class t , where $d \in (1, z)$. X_{ij} represents the value of the feature V_j in the training sample X_i .

In the same feature dimension of different objects, the feature values are not irrelevant, and they contain more or less-dependent relationships. The nonindependent and identically distributed relationship within the feature in this paper refers to the relationship within the feature between different objects and the coupling within the feature is reflected in the feature value. In addition, the nonindependent and identically distributed relationship between features and categories, that is, the class attribute feature α_{ry} , is considered to be incorporated.

Definition 5. We set V_j^Y and $V_j^{X_i}$ to be the eigenvalues of Y and X on the feature V , respectively. Among them, X is the sample in the training set belonging to t category, Y is the instance to be tested, and $t \in [1, z]$. Then the intra-feature similarity $\delta_j^{la}(Y, X_i)$ between Y and X is the following:

$$\delta_j^{la}(Y, X_i) = a_{tj} \cdot (V_j^Y - V_j^{X_i})^{-2}. \quad (26)$$

Here, V_j represents a feature of the data set, the feature dimension is M , and $1 \leq j \leq M$.

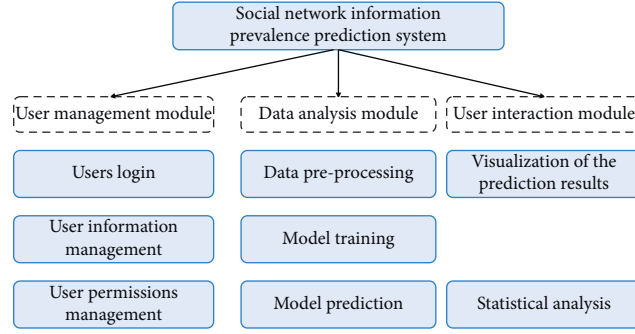


FIGURE 5: System function module diagram.

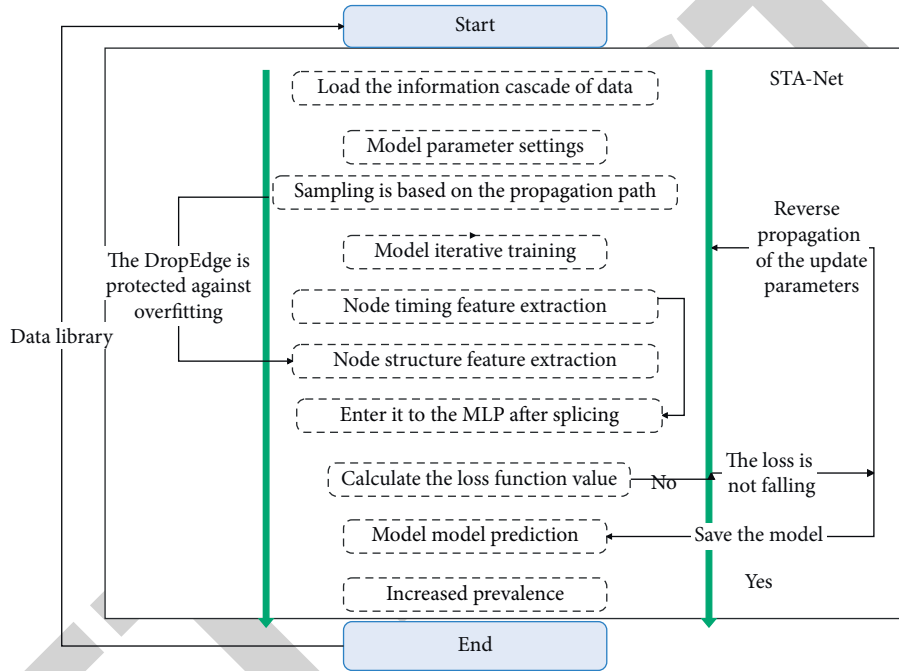


FIGURE 6: Information popularity prediction process.

Definition 6. We give the training set X , and X_i is the sample in the training set X . Then the interfeature similarity between feature R and feature D in training set X is

$$\delta_{hlj}^{Ie} = \frac{\sum_{X_i \in X} [(V_j^{X_i} - \bar{V}_j) \cdot (V_h^{X_i} - \bar{V}_h)]}{\sqrt{\sum_{X_i \in X} (V_j^{X_i} - \bar{V}_j)^2} \sqrt{\sum_{X_i \in X} (V_h^{X_i} - \bar{V}_h)^2}} \quad (27)$$

Here, \bar{V}_j and \bar{V}_h are the mean values of the features in the training set at V_j and V_h , respectively.

It can be seen from Definition 3 that the similarity between the features in the training set X is

$$\delta_j^{Ie} = \sum_{h=1, h \neq j}^M \delta_{j|h}^{Ie} \quad (28)$$

Definition 7. Under nonindependent and identical distribution, the feature correlation similarity $Csim(Y, X_i)$ between the sample to be tested and the training sample X in the data set is

$$Csim(Y, X_i) = \sum_{j=1}^M \delta_j^{Ia}(Y, X_i) \cdot \delta_j^{Ie} \quad (29)$$

Here, δ_j^{Ie} is the interfeature similarity of feature V in the training set X and δ_j^{Ia} is the intrafeature similarity between Y and X in the feature.

The flow chart of the improved CS_KNN algorithm is shown in Figure 4, and the specific flow is as illustrated in the Figure 4.

The input is as in sample $Y = \{Y_1, Y_2, \dots, Y_N\}$ to be tested, N is the total number of samples to be tested. We set up a training sample set $X = \{X_1, X_2, \dots, X_n\}$, and the total number of samples is n . The sample feature set is $A = \{A_1, \dots, A_j, \dots, A_M\}$, and the total number of features is M . The training sample category set is $C = \{C_1, \dots, C_t, \dots, C_z\}$, and the total number of categories is z , where $r = [1, z]$.

The output is as the predicted category C of the sample Y to be tested:

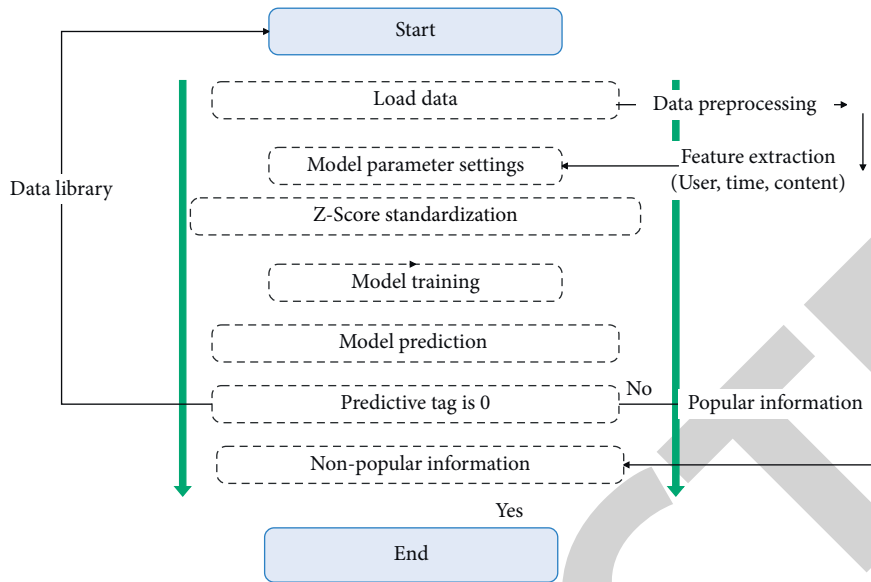


FIGURE 7: Popular information detection process.

Step1: it sets the initial value of parameter k , and the final value of this experiment is selected according to the experimental results.

Step2: it uses the introduced similarity δ^{Ia} within features and similarity δ^{Ie} between features to calculate the correlation similarity $Csim(Y, X_i)$ between the test instance Y and the training instance X .

Step3: it repeats the above process until all the training samples in the data set X have completed the correlation similarity calculation with the test instance Y .

Step4: at the end of the traversal, it sorts the most similar k tuples according to the degree of similarity, and then stores them in the set $T = \{T_1, T_2, \dots, T_k\}, f \in (1, k)$.

Step5: according to the decision-making method in the traditional KNN algorithm, it votes on the set of neighboring points T in the decision domain. The category with the largest value of the judgment function is used as the predicted category of the test instance Y , and the predicted category of the test instance is output.

4. Multimedia Network Information Popularity Prediction Model Based on Fast K-Nearest Neighbor Algorithm

This section mainly introduces the functional modules of the multimedia network information popularity prediction system. In general, the information popularity prediction system of the multimedia network mainly includes three modules: a user management module, a data analysis module, and a visual display module. The system function module diagram is shown in Figure 5.

The parameters of the predictive model include the learning rate, the number of model iterations, etc., which are initialized in a random manner. When the loss of 10 consecutive iterations in the verification set no longer decreases, the training ends, and the model progress and

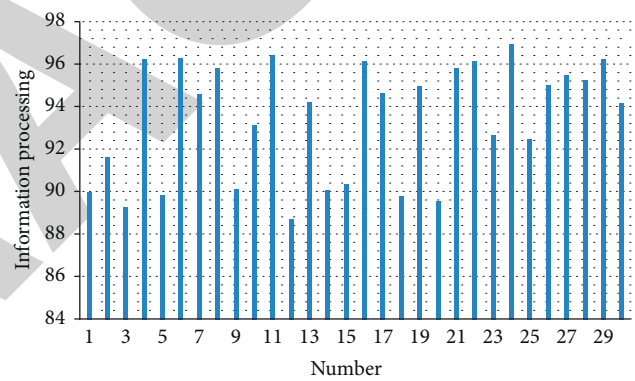


FIGURE 8: Statistical diagram of the information processing effect of the multimedia network information processing system based on the fast K neighbor algorithm.

corresponding indicators are visually monitored through TensorBoard(). Prediction is made by using the trained model, and the popularity prediction result is stored in the database to facilitate subsequent visual analysis and display. The information popularity prediction process is shown in Figure 6.

The overall process of popular information detection is shown in Figure 7.

This paper combines MATLAB to simulate the system proposed in this paper and obtains a large amount of data through the network to carry out experimental simulation processing to study the information processing effect of the system. The statistical test results are shown in Figure 8.

From the above research, we can see that the multimedia network information processing system based on the fast K neighbor algorithm proposed in this study has good multimedia network information processing functions. On this basis, this study evaluates the network information popularity prediction effect of the system and obtains the result shown in Figure 9.

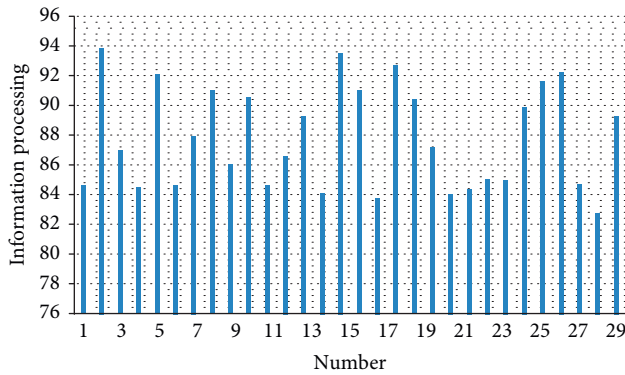


FIGURE 9: Statistical diagram of the prediction effect of the multimedia network information popularity prediction system based on the fast K neighbor algorithm.

From the above research, it can be seen that the prediction effect of the multimedia network information popularity prediction system based on the fast K neighbor algorithm proposed in this study is very good.

5. Conclusion

Multimedia data contain a large number of different types of information. However, this information is not static, but flowing, and has different manifestations in multiple scenarios. When a large number of social users spontaneously discuss a certain topic in a short period of time, information flow between users is formed. On the one hand, the data may contain malicious information such as rumors and false advertisements. On the other hand, most Internet applications are open, allowing data sharing and information exchange with other applications, thereby forming a positive feedback information flow effect. In-depth study of the laws of information flow in online media data is of great significance not only for accurately portraying the network information dissemination mechanism but also for group behavior analysis and social public opinion monitoring and avoiding the dissemination of malicious information. Moreover, it is of great significance to the theoretical research of sociology, administration, management, and other related disciplines. This article combines the fast K-proximity algorithm to predict and analyze the popularity of multimedia networks, constructs a corresponding intelligent model, and verifies the performance of the model to improve the effect of subsequent multimedia network information popularity prediction.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was sponsored by Research and Practice Project of Higher Education Teaching Reform in Henan Province and Research and Practice of Smart Classroom Teaching Effect Evaluation Based on Teaching Diagnosis Reform (No. 2019SJGLX791).

References

- [1] B. Y. Korniyenko and L. P. Galata, "Design and research of mathematical model for information security system in computer network," *Наукоємні технолоґії*, vol. 32, no. 2, pp. 114–118, 2017.
- [2] S. Chatterjee, A. K. Kar, and M. P. Gupta, "Alignment of IT authority and citizens of proposed Smart cities in India: system security and privacy perspective," *Global Journal of Flexible Systems Management*, vol. 19, no. 1, pp. 95–107, 2018.
- [3] A. Blagorazumov, P. Chernikov, G. Glukhov, A. Karapetyan, and L. Elisov, "The background to the development of the information system for aviation security oversight in Russia," *International Journal of Mechanical Engineering & Technology*, vol. 9, no. 11, pp. 341–350, 2018.
- [4] M. K. Özlen and I. Djedovic, "Online banking acceptance: the influence of perceived system security on perceived system quality," *Journal of Accounting and Management Information Systems*, vol. 16, no. 1, pp. 164–178, 2017.
- [5] S. Dotsenko, O. Illiashenko, S. Kamenskyi, and V. Kharchenko, "Integrated security management system for enterprises in industry 4.0," *Information and security : an international journal*, vol. 43, no. 3, pp. 294–304, 2019.
- [6] N. Y. Kim, S. Rathore, J. H. Ryu, H. P. Jin, and J. H. Park, "A survey on cyber physical system security for IoT: issues, challenges, threats, solutions," *Journal of Information Processing Systems*, vol. 14, no. 6, pp. 1361–1384, 2018.
- [7] S. U. Hani and A. T. Alam, "Software development for information system - achieving optimum quality with security," *International Journal of Information System Modeling and Design*, vol. 8, no. 4, pp. 1–20, 2017.
- [8] M. Sun, I. Konstantelos, and G. Strbac, "A deep learning-based feature extraction framework for system security assessment," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5007–5020, 2019.
- [9] K.-K. R. Choo, M. M. Kermani, R. Azarderakhsh, and M. Govindarasu, "Emerging embedded and cyber physical system security challenges and innovations," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 3, pp. 235–236, 2017.
- [10] D. Li, Z. Cai, L. Deng, X. Yao, and H. H. Wang, "Information security model of block chain based on intrusion sensing in the IoT environment," *Cluster Computing*, vol. 22, no. S1, pp. 451–468, 2019.
- [11] A. Safi, "Improving the security of internet of things using encryption algorithms," *International Journal of Computer and Information Engineering*, vol. 11, no. 5, pp. 558–561, 2017.
- [12] W. A. Cram, J. D'arcy, J. D'Arcy, and J. G. Proudfoot, "Seeing the forest and the trees: a meta-analysis of the antecedents to information security policy compliance," *MIS Quarterly*, vol. 43, no. 2, pp. 525–554, 2019.
- [13] V. H. Le, V. O. Phung, and N. H. Nguyen, "Information security risk management by a holistic approach: a case study for Vietnamese e-government," *IJCSNS International Journal of Computer Science and Network Security*, vol. 20, no. 6, pp. 72–82, 2020.

- [14] A. B. Lopez, K. Vatanparvar, A. P. Deb Nath, S. Yang, S. Bhunia, and M. A. Al Faruque, "A security perspective on battery systems of the internet of things," *Journal of Hardware and Systems Security*, vol. 1, no. 2, pp. 188–199, 2017.
- [15] P. B. Lowry, T. Dinev, and R. Willison, "Why security and privacy research lies at the centre of the information systems (IS) artefact: proposing a bold research agenda," *European Journal of Information Systems*, vol. 26, no. 6, pp. 546–563, 2017.
- [16] S. Trang and B. Brendel, "A meta-analysis of deterrence theory in information security policy compliance research," *Information Systems Frontiers*, vol. 21, no. 6, pp. 1265–1284, 2019.
- [17] M. Rajesh, "A signature based information security system for vitality proficient information accumulation in wireless sensor systems," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 9, pp. 367–387, 2018.
- [18] O. Na, L. W. Park, H. Yu, Y. Kim, and H. Chang, "The rating model of corporate information for economic security activities," *Security Journal*, vol. 32, no. 4, pp. 435–456, 2019.
- [19] K. Kavitha and R. Neela, "Optimal allocation of multi-type FACTS devices and its effect in enhancing system security using BBO, WIPSO & PSO," *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 777–793, 2018.
- [20] Z. Turskis, N. Goranin, A. Nurusheva, and S. Boranbayev, "Information security risk assessment in critical infrastructure: a hybrid mcdm approach," *Informatica*, vol. 30, no. 1, pp. 187–211, 2019.