**INTERNATIONAL JOURNAL OF RESEARCH – GRANTHAALAYAH**

**A knowledge Repository**

Science

# INFORMATION SECURITY ASSESSMENT BASED ON MACHINE LEARNING TECHNOLOGY-FUZZY-GRA-AHP

**Mei-Er Zhuang [1], Wen-Tsao Pan [*1], Jia-Yan Zhou [1], Qian Yang [1], Wan-Ting Hong [1]**
[1] School of Business, Guangdong University of Foreign Studies, Guangzhou, China

**Abstract**

With the advent of the information age, information security has become an urgent problem to be solved. Various application and platforms have not only brought convenience to people, but also brought hidden dangers - information security risks. This paper uses some of the machine learning technology - fuzzy computing and gray relation analysis (GRA), to analyze data of the three major video platforms of China, and takes the information security level as a new criterion to conduct the evaluation of their performance. An assessment model is constructed based on machine learning technology, namely the combination of fuzzy computing and GRA and analytic hierarchy process (AHP). Conclusions can be drawn as follows. First, consumers' perception of video platform information security level is constantly being strengthened. Second, information security risks are affecting consumers' choice decisions about video platforms, and the weights will continue to increase. Third, video platforms are paying more attention to information security construction.

*Keywords:* Information Security; Machine Learning; Fuzzy Computing; Gray Relation Analytic (GRA); Analytic Hierarchy Process (AHP).

## 1. Introduction

The number of Internet users in China reached 829 million, among which the number of online video users reached 612 million, accounting for 73.9% of the total netizens. Jiang Liqiu (2015) believes that the development focus of the online video industry has changed from "content is king" to "platform is king", that is, while providing high-quality content, attention should be paid to the launch of a series of initiative service such as interactive feedback and personalized recommendation. However, there is a kind of service, namely information security, which is often ignored by those online video platforms. They failed to protect the privacy of users, which led to security risks such as the leak of personal information or financial losses.

In China, the three major platforms, iQiyi, Tencent Video and Youku, have concentrated most users. In terms of user scale, the users of the three major platforms account for nearly 90% of the total industry users; in terms of content and traffic, about 80% of the newly launched self-produced programs in 2018 are broadcast on these three platforms, and the broadcast volume accounts for more than 80% of the total broadcast volume.

However, there is few study of the performance evaluation of information security level of video platforms. The existing literature rarely mentions how the user make the final decision in selecting a video platform under the consideration of information security. Actually, information security has become an extremely important issue in the society.

Previous research mostly focused on the characteristics and functions of video without paying attention to the factor of information security. Sun Yonglu (2014) collects content evaluation, Douban Index, Alexa ranking, content popularity score and video quality, and uses big data crawler technology and fuzzy comprehensive evaluation method to find out how the user makes final decisions when selecting the video platform. Shen Junwei (2015) constructed the core competitiveness evaluation model of the video platform through two levels of indicators: content scarcity and user experience. The evaluation values were obtained according to the grey correlation method, and the core competitiveness of each platform was sorted. But these papers neglected the influence of information security when users select the video platforms. As for the research of information security of video, many scholars have studied this field from different angles. Jiang Chengming et al. (2012) designed an online video classification method based on multimodal features and thereby the security supervision of the videos can be realized. This method filters the input videos by such different features as audio, color motion and space-time features in a specific order. Chen Yingjie (2009) introduced the overseas practices in information security management of some online services such as community applications, information publishing applications and information retrieval applications. Lu xin et al. (2019) pointed out that effective measures should be taken to intercept and restore video signals, and a complete video information interception system should be established to collect, transmit, analyze and perceive video signals, which can effectively prevent video information from leaking. Wu Youxin et al. (2009) believe that online video can be used to express and present self, share information, and develop interpersonal relationships. At the same time, negative effects and problems such as network infringement, network leakage, cyber violence, and network rumors have followed. These problems make the country's control over the Internet more difficult, and the situation facing national information security is more complicated. Information security issues have become an important potential threat to national security.

In order to explore the influence of information security level to user's choice decision about video platforms, we take machine learning technology as the research method and construct an assessment model based on fuzzy computing, grey relation analysis (GRA) and analytic hierarchy process (AHP). In previous studies, some of the methods mentioned are used to analyze many problems effectively, such as evaluating the environmental internal control system of universities by Fuzzy-AHP (2017) or the operation modes of two typical chemical logistics enterprises (2015).In terms of indicator selection, the primary and secondary indicators of this paper refer to the research results of Xue Song (2017) on the competitiveness of video websites. It is worth noting that, because the self-made content has been verified for the "back-feeding" effect of the video

platform (Chai Xiaotian, 2017), this article also adds "home-made content" as the secondary indicator of the first-level indicator "content".

In brief, to make further research in this field, this paper uses machine learning technology, Fuzzy-GRA-AHP, to assess how the information security level of different video platform affect the users' decision when facing these alternatives.

## 2. Research Methods

### 2.1. Web Crawler Technology

Web crawler automatically crawls programs or scripts of World Wide Web information according to certain rules to obtain and update the content and retrieval methods of Internet search engines or other similar websites. The data is usually captured from the official website, and there are two types of target data to ensure the accuracy and reliability of the data: one is large-scale data that approximates fixed features in the webpage, and the other is the data of the webpage with fixed URL and real-time information update (Shu Wanchang, 2018).With the increasing competition in the network video industry and the arrival of the big data era, data collection is crucial for the update of the video platform. As far as academic research is concerned, using crawler technology to crawl valuable data is the best choice to make up for the scarcity of its own data.

### 2.2. Fuzzy Theory

Fuzzy theory refers to the theory that uses the basic concept of fuzzy sets or continuous membership functions. Based on fuzzy mathematics, qualitative evaluation can be transformed into quantitative evaluation, that is, an overall evaluation of things or objects subject to various factors.

**Step 1 Determine the Set of Factors for the Evaluation Object**
Let $U = \{u_1, u_2, \cdots, u_m\}$ be a set of m kinds of evaluation indexes to be evaluated, where: m is the number of evaluation indexes.

**Step 2 Determine the Comment Set of the Evaluation Object**
Let $V = \{v_1, v_2, \cdots, v_n\}$ be a set of evaluation levels composed of the evaluation results that the evaluator may make on the evaluated object, where: v_j represents the jth evaluation result, and n is the total evaluation result number,it is divided into 5 levels in this article.

**Step 3 Determining the Weight Vector of the Evaluation Factor**
In general, the factors affecting the evaluated things are inconsistent, so the weight distribution of the factors is a fuzzy vector on the comment set, which is recorded as $A = (a_1, a_2, \cdots, a_m)$, where: $a_m$ is the first weight corresponding to the m factors $X_m$, and satisfies $a_1 + a_2 + \cdots + a_m = 1$.

**Step 4 Perform Single-Factor Fuzzy Evaluation and Establish Fuzzy Relation Matrix R**
The evaluation is made from one factor alone to determine the degree of membership of the evaluation object to the evaluation set V, which is called single factor fuzzy evaluation.

After constructing the hierarchical fuzzy subset, the evaluated objects are quantized m from each factor one by one, that is, the membership degree of the evaluated object to each level of fuzzy subset is determined, and then the fuzzy relation matrix is obtained:

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & & r_{2n} \\ & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{pmatrix}$$

Where: $r_i$ is the one-factor evaluation set of the index $u_i$ that is, $r_i$ is the fuzzy subset on the comment set V, denoted as $r_i = (r_{i1}, r_{i2}, \cdots, r_{in})$; $r_{ij}$ represents an evaluated object from the factor $u_i$ The membership of the hierarchical fuzzy subset $v_j$, and $\sum r_{ij} = 1$.

**Step 5 Comprehensive Evaluation**
For the weight distribution $A = (a_1, a_2, \cdots, a_m)$ a comprehensive evaluation $B = A \cdot R = (b_1, b_2, \cdots, b_m)$ is obtained. If $B = \max\{b_1, b_2, \cdots, b_m\}$, according to the principle of maximum membership degree, the comprehensive evaluation result is to evaluate the thing B.

## 2.3. Grey Relation Analysis (GRA)

The grey system theory puts forward the concept of grey relation analysis for each subsystem and intends to seek the numerical relationship between subsystems (or factors) in the system through certain methods. A measure of the magnitude of the association between two systems, which varies over time or from different objects, is called the degree of association.

**Step 1 Identify Reference Series that Reflect System Behavior Characteristics and Comparison Series that Affect System Behavior**
A sequence of data that reflects the behavioral characteristics of the system, is called the reference sequence. A sequence of data that is a component of factors that affect system behavior is called a comparison sequence.

Let the reference sequence be $Y = \{Y(k)|k = 1,2,\wedge, n\}$, and compare the sequence to $\{X_i(k)|k = 1,2,\wedge, n\}, i = 1,2,\wedge, m$.

**Step 2 Non-Dimensionalization of Reference Series and Comparison Series**
Because the physical meanings of various factors in the system are different, the dimensions of the data are not necessarily the same, it is not convenient to compare, or it is difficult to get a correct conclusion when comparing. Therefore, when performing grey relation analysis, it is generally necessary to perform dimensionless data processing:

$$x_i(k) = \frac{X_i(K)}{X_i(1)}$$

**Step 3 Find the Grey Correlation Coefficient (Xi) of the Reference Series and Comparison Series**

The degree of association is essentially the degree of difference in geometry between curves. Therefore, the difference between the curves can be used as a measure of the degree of association. For a reference sequence X0, there are several comparison sequences X1, X2, ..., Xn, and the correlation coefficient ξ(Xi) of each comparison sequence and the reference sequence at each time (i.e. each point in the curve) can be calculated by the following formula:

$$\xi_{0i} = \frac{\Delta\,(min) + \rho\,\Delta\,(max)}{\Delta_{0i}\,(k) + \rho\,\Delta\,(max)}$$

Where: Δmin is the second-order minimum difference, Δmax is the two-stage maximum difference, and $\Delta_{0i}(k)$ is the absolute difference between each point on the comparison series Xi curve and each point on the reference sequence X0 curve.

ρ is the resolution coefficient, and the smaller ρ is, the larger the resolution is. Generally, the value range of ρ is (0, 1), and the specific value may depend on the situation. When $\rho \leq 0.5463$, the resolution is the best, usually $\rho = 0.5$.

**Step 4 Relevance**

The correlation coefficient is the correlation degree between the comparison sequence and the reference sequence at each moment (i.e. each point in the curve), so its number is more than one, and the information is too scattered to facilitate the overall comparison, so it is necessary to The correlation coefficient of each point in the curve is concentrated into one value, that is, the average value is obtained as a quantity representation of the degree of correlation between the comparison series and the reference sequence. The correlation formula is as follows:

$$r_i = \frac{1}{N} \sum_{k=1}^{N} \xi_i(k)$$

The closer the ri value is to 1, the better the correlation.

**Step 5 Ranking of Relevance**

The degree of association between factors is mainly described by the order of relevance, not just the degree of relevance. If r1 < r2, the reference sequence Y is more similar to the comparison sequence X2.

After calculating the correlation coefficient between the Xi(k) sequence and the Y(k) sequence, the average value of each type of correlation coefficient is calculated, and the average value ri is called the degree of association between Y(k) and Xi(k).

### 2.4. Analytic Hierarchy Process (AHP)

Analytic Hierarchy (AHP) refers to the decomposing of elements that are always related to decision-making into goals, criteria, and programs. On this basis, qualitative and quantitative analysis methods are used.

**Step 1 Establishing an Indicator Hierarchy**
Layer the factors involved in the problem: the target layer (the purpose of the problem); the criteria layer (choose the various measures to be taken to achieve the overall goal, the guidelines that must be followed); the solution layer (used to solve the problem) Measures, programs, etc.). Put the various factors to be considered into the appropriate levels, and draw a hierarchy diagram to clearly express the relationship of these factors.

**Step 2 Building a Pairwise Comparison Judgment Matrix**
When comparing the importance of the ith element to the jth element relative to a factor above, it is described using a quantified relative weight $a_{ij}$. Let a total of n elements participate in the comparison, then $A = \left(a_{ij}\right)_{n \times n}$ is called a pairwise comparison matrix. The value of $a_{ij}$ in the pairwise comparison matrix can be referred to Satty's proposal and assigned according to the following scale. $a_{ij}$ takes the value between 1-9 and its reciprocal.

**Step 3 Hierarchical Single Sort**
Hierarchical single sorting refers to the ordering of the importance of each factor of this level for a factor above. The feature vector corresponding to the largest eigenvalue of the judgment matrix is normalized (so that the sum of the elements in the vector is equal to 1), and is denoted as W. The maximum eigenvalue $\lambda_{max}$ is calculated according to the formula $A \cdot W = \lambda_{max} \cdot W$.

**Step 4 Consistency Test**
Since λ is continuously dependent, the more λ is larger than n, the more the inconsistency of A is. The consistency indicator is calculated in CI

$$CI = \frac{\lambda_{max}(A) - n}{n - 1}$$

In general, CI=0, there is complete consistency; the larger the CI, the more serious the inconsistency of A.

Considering that the deviation of consistency may be caused by random reasons, when testing whether the judgment matrix has satisfactory consistency, it is also necessary to compare the CI with the random consistency indicator RI (which can be obtained by looking up the table). Test coefficient CR:

$$CR = \frac{CI}{RI}$$

When CR < 0.1, it is determined that the paired comparison array A has satisfactory consistency, or the degree of inconsistency is acceptable; otherwise, the pairwise comparison matrix A needs to be adjusted until satisfactory consistency is achieved.

**Step 5 Hierarchical Total Ordering**
The process of ranking weights that determines the relative importance of all factors of a layer to the overall goal is called the hierarchical total ordering. This process is carried out in order from the highest level to the bottom. For the highest level, the result of sorting the hierarchical order is the result of the total sorting.

### 3. Empirical Analysis

### 3.1. Data Source

Through machine learning and big data mining technology, users' comments on the video playback platform on the Internet are collected and used to conduct corresponding analysis. These comments are about three major video platforms in China, iQiyi video, Tencent video and Youku video. The text evaluation of the video playback platform is converted into quantifiable values as the raw data of the secondary indicators of the criteria layer in the empirical analysis.

There are three primary indicators in the criteria layer, namely content (B1), brand (B2), and function (B3), which are calculated by sub-indicators. The information security level is considered and included as one of the sub-indicators of brand(B2), which is one of the most important innovations and contributions of this research.

### 3.2. GRA Based on Fuzzy Computing

The concrete steps of grey relational analysis based on fuzzy calculation are as follows.

**Step 1**: Using R console, the secondary indicators are fuzzed and the individual evaluation values of each user for the primary indicators are calculated. 12 secondary indicators should be classified and merged according to the degree of correlation among variables. In this way, users' individual evaluation data of three video platforms on three primary indicators can be obtained.

**Step 2**: Calculate the overall primary indicators of each enterprise by secondary fuzzy analysis of the individual primary indicators data. Since after the first round of fuzzy calculation, there are still 63 data for each scheme's primary index value, and the AHP analysis requires only one final value for each scheme's primary index, so the second round of grey correlation analysis is needed. In the first step, the individual evaluation data of the primary indicators are calculated again by using the fuzzy method, and the correlation degree is calculated as the final value of each primary indicator. The results are shown in table 1.

Table 1: GRA Result Based on Fuzzy Computing

| Relational degree | C1 | C2 | C3 |
|---|---|---|---|
| B1 | 0.590771 | 0.634319 | 0.642216 |
| B2 | 0.93684 | 0.827892 | 0.763432 |
| B3 | 0.872093 | 0.596016 | 0.698267 |

### 3.3. Assessment Model Based on AHP

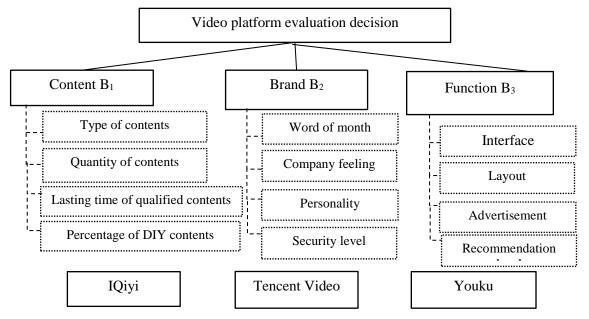According to principle of AHP, an empirical analysis model is constructed in figure 1.



Figure 1:  Model Based on Fuzzy-GRA-AHP

According to the calculated correlation degree and the principle of AHP, the weights of mutual factors of criterion layer B and the weights of scheme layer C to each criterion layer B are calculated, as shown in the following two tables. (see table 2 and 3)

Table 2: Criterion Layer B Mutual Factor Weight

| A | B1 | B2 | B3 |
|---|---|---|---|
| B1 | 1 | 0.583555 | 0.757873 |
| B2 | 1.713635 | 1 | 1.298718 |
| B3 | 1.319483 | 0.76999 | 1 |

Table 3: Scheme Layer to the Criterion Layer Mutual Factor

| A | B1 | | | B2 | | | B3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 |
| C1 | 1 | 0.630600 | 0.677417 | 1 | 0.766185 | 1.064265 | 1 | 0.841222 | 0.919729 |
| C2 | 1.585792 | 1 | 1.074243 | 1.305167 | 1 | 1.389044 | 1.188746 | 1 | 1.093325 |
| C3 | 1.476195 | 0.930888 | 1 | 0.939616 | 0.719920 | 1 | 1.087277 | 0.914641 | 1 |

The criterion layer mutual factor weight value and the program layer to the criterion layer mutual factor weight value are substituted into the AHP program, and the R software is used to calculate the plan layer single order weight and the total order weight. (see table 4)

Table 4: Result of Fuzzy-GRA-AHP

| Criteria layer B | | Content B1 | Brand B2 | Function B3 | Total sort weight |
|---|---|---|---|---|---|
| Criteria layer weight value | | 0.2479471 | 0.4248910 | 0.3271619 | |
| Scheme layer single sort weight | IQiyi C1 | 0.2463550 | 0.3080597 | 0.3052482 | 0.2918404 |
| | Tencent Vedio C2 | 0.3901268 | 0.4025266 | 0.3628627 | 0.3864756 |
| | Youku C3 | 0.3635182 | 0.2894137 | 0.3318892 | 0.3216840 |
| Value of cr | | 0.0009463703 | | | |

From the data results above, it can be seen in the criterion layer that the weight of content is 0.2479471, which is the lowest among the three indicators, while the weight of brand is 0.4248910, which is the highest among the three primary indicators, while the weight of function is 0.3271619, whose weight is between the other two primary indicators.

According to the analysis of fuzzy-GRA-AHP, in terms of the total sort weight, C1 < C2 < C3. In other words, Tencent Video performs best and is user's first choice or best decision. Youku ranks second, as the user's secondary choice, while iQiyi ranks last among the three video sites, which means user's choice preference is the lowest. The research shows that information security is of great importance in the performance of online video platforms. The reason why Tencent video ranks first in the study is that its perceived value of its brand is much higher that those of its competitors, and one of the factors that exerts positive effects on its brand is the information security level. Therefore, the higher level of information security is, the more trustworthy a video platform will be. Thus, more users have a preference for this video platforms, accounting for the best performance of Tencent video in the assessment of this paper.

The CR value of the test coefficient in AHP analysis is 0.0009463703 < 0.1, so it passes the consistency test, which shows that the analysis meets the requirement of consistency.

## 4. Conclusions and Recommendations

Since the awareness of information security is being improved rapidly, video platforms had better pay more attention to information security construction while developing. Due to the explosive growth of OTT usage in China, e-Marketer has increased the market share of Tencent Video, iQiyi and Youku. In 2018, 24% of Chinese online TV viewers that will exceed 29% by 2020 subscribe to Tencent Video.

Investing in content has been a key theme for platforms from BAT (Baidu, Alibaba and Tencent). Youku from Alibaba achieved the highest content expenditure growth in 2018 and received live broadcast rights for the FIFA World Cup this year. Therefore, it is expected that Youku's user base will grow by 55% this year, and by the end of 2019, the share of its online video market will exceed the second-ranked iQiyi.

At present, the market size of the video platform exceeds 7 billion. The focus of competition is gradually transformed into users by resources and IP. As an important factor affecting user choice, the brand has gradually been valued by major video sharing websites that have launched their own brand upgrading strategy. Since the beginning, iQiyi has focused on its content. In 2015, due to Alibaba's sponsorship, the management team of Youku changed. In the case that the running-in period has not yet passed and the person in charge is not familiar with the entertainment industry,

its brand strategy tends to be conservative rather than innovative. The traditional play is difficult to meet the new needs of new users, resulting in rapid loss of users and rapid decline in word of mouth. In a short period of two years, Youku was overtaken by iQiyi and Tencent Video.

In terms of functional experience, the three major video platforms, regardless of interface layout, guide window, and number of advertisements, are basically the same, verifying the serious homogenization of the network video industry. In order to attract more users, the three major video platforms have adjusted their respective development priorities. In order to cater to young people's tastes of interaction, fun and sociality, iQiyi has introduced functions such as game centers, stores and friends, and gradually shown its development towards the entertainment industry chain. Tencent Video focuses on enhancing the social effects of fans and opening a film and television circle. Live, rice ball and other sections. Youku focuses on the UGC platform, user personalized subscriptions, and the need for users to watch live broadcasts.

It can be predicted that the competition of the network video industry in the future will be a comprehensive confrontation of content, brand and function. The pursuit of differentiation will become the key development strategy of major video platforms, but the real differentiation is difficult to achieve. The core of the UCG pattern lies in making users really participate in the content output and value creation of the video platform, in other words, through the video carrier, the video providers and viewers are aggregated in different categories, triggering their participation, discussion and interaction, and maximizing user stickiness. Therefore, the future development direction of video platforms would be to attract users with content, precipitate users with social relationships, grasp the wave of the era of mobile big data, integrate various resources across borders, and make their own characteristics. Overall, video platforms should enhance the customer-perceived value of their own content, function and brand and especially pay more attention to one of most significant sub-indicators of brand, namely information security level. Optimizing the information security system is conducive to a more trustworthy image of the brand and improvement of their performance.

## References

[1] Jiang Liqiu. Research on Video Website Communication Mode and Development Strategy. [D], Hunan Normal University,2015
[2] Xie Xinzhou, Huang Yuting. The application of big data in network video needs more rationality. Visit Ge Chengzhi, Dean of Academy of IQYI Data. [J]. News and Writing, 2018 (09): 82-85
[3] Zhao Meng. The Construction of Internal Control Evaluation System in University based on Environmental Target—Using AHP-Fuzzy Analysis Method. Communication of Finance and Accounting[J].2017, (34),55-57+4
[4] Chen Hongyan. Fuzzy Analysis of Operational Safety Evaluation of Chemical Logistics Enterprises. Journal of Commercial Economics.2015(11),25-26
[5] Xue Song. Research on the Competitiveness of Video Websites Based on Multiple Regression [D]. Yanshan University, 2017
[6] Chai Xiaotian. Research on the "Feedback" Effect of Self-made Programs on TV under the Background of Media Integration [D]. Zhengzhou University, 2017.
[7] Sun Yonglu. Research on Automatic Discovery and Evaluation Method of Internet Video Service Website [D]. China University of Science and Technology, 2014.
[8] Shen Junwei. Research on Core Competence of Internet Video Website [J]. Times Finance, 2015 (11): 208-209.

[9] Jiang Chengming, Jiang Xinghao, Sun Tanfeng. Video Filtration for Content Security based on Multimodal Features[J]. Information Security and Communications Privacy.2012 (3):76-77

[10] Chen Yingjie. Overseas Practices in Information Security Policies Concerning Internet New Applications[J]. Modern Science & Technology of Telecommunications.2009(5):37-41

[11] Lu Xin, Zhang Meng, Tan Yueyang. Research on Signal Processing and Reduction Technology of Computer Video Leakage Information[J]. Science and Technology Innovation Herald.2019(6):142-143

[12] Wu Youxin, Wu Bo, Zhang Ming. Information Security Hidden Troubles in Network Video and Countermeasures.China Radio & TV Academic Journal[J].2009(11):44-45

[13] Shu Wanchang. Application and Analysis of Reptilian Technology in Big Data Field [J]. Scientific and Technological Innovation, 2018 (36): 91-92.

[14] 199 IT Internet Data Center.www.emarketer.com. 2018,7

*Corresponding author.
*E-mail address:* teacherp0162@ 126.com