Academic Journal of Business & Management
ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

Research on Self-media Original Video Protection Based on Machine Learning

——Based on the original author's perspective

Fan Wu[1], Wenxin Guan[2], Shouyu Wei[3]

[1] *School of Applied Mathematics, Anhui University of Finance and Economics*
[2] *School of Business Administration, Anhui University of Finance and Economics*
[3] *School of Finance, Anhui University of Finance and Economics*

ABSTRACT. *In order to understand the willingness of self-media video users to protect the copyright of original videos and the improvement direction of the self-media video ecosystem, 1,500 questionnaires were actually distributed to users on the entire network video platform as survey objects to obtain sample data, and the rights of authors were affected by machine learning. In-depth analysis of the factors of will, we mainly found that although the benefits obtained by plagiarists and the cost of video production can most stimulate the author's willingness to defend rights, this is not controllable. Therefore, we believe that reducing the cost of rights protection (the second most important) is the most feasible direction to increase the author's willingness to protect rights. In addition, it is also found that the support of the platform and the report of the audience will have a direct impact on all factors, and the feedback factor and cost factor will have a direct impact on the induction factor. Finally, we made corresponding suggestions.*

KEYWORDS: *Self-media video ecosystem; rights protection; machine learning; Python*

1. Introduction

With the frequent occurrence of online original video infringement incidents, the protection of original culture is imminent, and different scholars have given suggestions from different perspectives. Zhu Chen [1] (2020) made recommendations from three aspects. Cong Lixian [2] (2019) mainly gives suggestions from the perspective of the platform, but there are few literatures considered from the perspective of numerical quantification. Therefore, this article will make up for the weaker aspects of the current research and propose to further improve the rights protection of original authors of self-media videos method.

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

2. An Analysis of the Differences of Self-media Original Author Information Features on Rights Protection

This module aims to study the influence of the creative environment factors of the original authors of the media (in the field of video) on the willingness to protect rights. The author's willingness to defend rights is the extent to which the original author of the media is willing to take measures to defend rights after discovering that the published work has been plagiarized. It is mainly divided into unwillingness, less willingness, general willingness, greater willingness, and strong willingness. The creators of different video partitions have been in different types of environments for a long time, and this type may affect the author's cognition subtly, and then have an impact on the rights protection. Therefore, we use contingency tables to analyze the influencing factors of creators in different fields to protect their rights.

2.1 Contingency table analysis of video categories & rights protection

We have produced a cross-tabulation and inspection results of video categories and rights protection awareness, as shown in Table1and Table2.

Table1 Cross-tabulation of video categories & rights protection awareness

category	Rights consciousness					total
	Willingness is small	Less willing	Average will	Willing	Very willing	
Anime	2	4	5	6	8	25
	9.52%	14.29%	19.05%	23.81%	33.30%	100%
Music	3	6	4	2	1	16
	18.18%	36.36%	27.28%	13.63%	4.55%	100%
Games	7	12	4	5	2	30
	23.08%	38.46%	12.82%	17.95%	7.69%	100%
Technology	3	6	17	7	6	39
	7.32%	14.63%	43.90%	17.07%	15.38%	100%
Digital	2	7	14	11	7	41
	4.08%	16.33%	34.69%	26.53%	17.07%	100%
Lifestyle	15	22	20	6	2	65
	22.78%	34.18%	30.38%	8.86%	3.80%	100%
Ghosts	16	15	11	6	2	50
	32.31%	29.23%	21.54%	12.31%	4.62%	100%
Fashion	12	8	10	5	3	38
	32.20%	20.34%	25.42%	13.56%	8.47%	100%
Advertising	1	2	3	1	1	8
	16.67%	25.00%	33.33%	16.67%	8.33%	100%
Film and television	12	10	13	4	6	45
	26.92%	23.08%	28.85%	9.62%	11.54%	100%
Learning	1	3	7	12	5	28
	2.78%	11.11%	25.00%	44.44%	16.67%	100%
other	6	11	7	8	5	37
	17.14%	28.57%	20.00%	22.86%	11.43%	100%

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

Table2 Chi-square test of video category and rights protection

project	Statistics	df	Progressive Sig. (both sides)
Pearson chi-square	53.6564.	17	0.000
Likelihood ratio	41.2354	17	0.000
Linear and nonlinear combination	1.851	1	0.174
N in valid cases	422	—	—

It can be seen from Table2 that the chi-square P value = 0.000<0.05, it can be seen that there are obvious differences in the rights protection rights of authors of different video categories. The specific differences are visually analyzed through the contingency Table1.

It can be seen from Table1 that the authors engaged in the production of animation videos have the highest proportion of options that have a high willingness to protect their rights. This may be due to the generally high cost of animation video production, and the great effort of the author, so it will be particularly sensitive to the protection of the video. We recorded the production cost options for original videos in the questionnaires of this category as 1, 2, 3, 4, and 5. By solving the average, we found that the average option value for this category of authors is 4.76, which is much higher than the overall average. Confirmed our thoughts. In the options with greater willingness, it can be seen that the selection ratio of the authors who make learning videos is extremely high, which is much higher than other categories. At the same time, the sum of the willingness of the learning area and the option of greater willingness is the sum of all video categories. The highest, which shows that the producers of learning videos have a much higher awareness of plagiarism than other categories, and the learning atmosphere will have a subtle influence on cognition. Among the options with general will, the video author option of science and technology accounts for the highest proportion, and such authors express a neutral attitude towards rights protection. Among the less willing options, the producers of game videos accounted for the highest percentage of such options. Among the options with little intention, the authors of ghost and animal videos are at the top of the proportion of these options, which shows that these two types of authors generally lack awareness of rights protection. From the cross table 5-1, we can see that authors' awareness of rights protection in different video fields will be obviously different. This provides ideas for us to take targeted protective measures later.

3. Analysis of Factors Influencing Self-media Original Authors' Rights Protection

In order to further study the influencing factors of authors' rights protection consciousness, we first use random forest to rank the importance of scale indicators (first-level indicators) and carefully understand the importance of these indicators. Then the factor analysis method is used to reduce the dimensionality of the

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

first-level indicators to obtain the main factors (second-level indicators) that affect the author's rights protection awareness. Finally, in order to explore the interaction between variables, we use the pls-pm model to discuss the secondary indicators.

3.1 Analysis of the importance of factors based on random forest

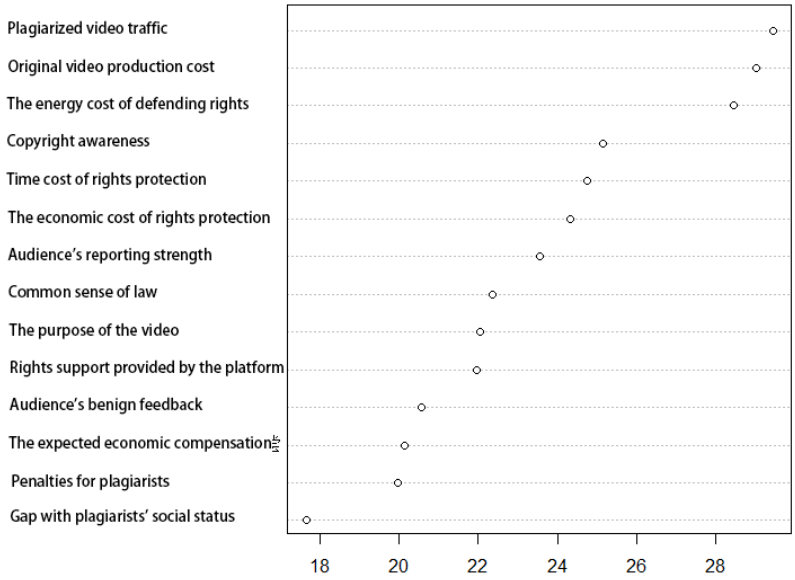

Figure1 Ranking diagram of various influencing factors of the right to protect rights

As can be seen from Figure1, the top priority is the traffic obtained from plagiarized video, the cost of original video production, and the cost of rights protection.

3.2 Factor analysis that affects original authors' rights protection

In this part, through factor analysis of the factors that affect the original author's rights protection awareness, these factors are simplified and reduced in dimension. We converted the 14 measurement indicators into mean scores, using factor load as the weight, condensed into 5 factor influence degree indexes, and obtained the factor weight. The results are detailed in Table 3 and Table 4.

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

Table3 Factor rotation orthogonal factor table

index	factor					Secondary indicators
	1	2	3	4	5	
Original video production cost	0.883	0.226	0.121	0.147	0.171	Inducible factor
The purpose of video plagiarism	0.842	0.136	0.244	0.135	0.097	
Plagiarized video traffic	0.791	0.221	0.124	0.086	0.113	
Audience's reporting strength	0.127	0.841	0.115	0.134	0.315	Environmental Factors
Rights support provided by the platform	0.012	0.801	0.322	0.089	0.211	
Expected economic compensation	0.012	0.352	0.813	0.126	0.141	Feedback factor
Benign feedback	0.089	0.348	0.791	0.123	0.166	
Penalties for plagiarism	0.105	0.136	0.679	0.213	0.180	
Time cost of rights protection	0.230	0.344	0.111	0.802	0.192	Cost factor
The economic cost of rights protection	0.154	0.188	0.029	0.797	0.207	
The energy cost of defending rights	0.173	0.348	0.136	0.744	0.129	
Gap with plagiarists' social status	0.038	0.021	-0.041	0.669	0.034	
Copyright awareness	0.135	0.181	0.285	0.061	0.721	Quality factor
Common sense of law	0.236	0.411	0.113	0.125	0.633	

Table 4 Statistical table of influencing factors of author's willingness to defend rights

index	Factor load	Weights	name
Original video production cost	0.883	0.351	Inducible factor
The purpose of video plagiarism	0.842	0.335	
Plagiarized video traffic	0.791	0.314	
Audience's reporting strength	0.841	0.512	Environmental Factors
Rights support provided by the platform	0.801	0.488	
Expected economic compensation	0.813	0.356	Feedback factor
Benign feedback	0.791	0.346	
Penalties for plagiarism	0.679	0.297	
Time cost of rights protection	0.802	0.266	Cost factor
The economic cost of rights protection	0.797	0.265	
The energy cost of defending rights	0.744	0.247	
Gap with plagiarists' social status	0.669	0.222	
Copyright awareness	0.721	0.532	Quality factor
Common sense of law	0.633	0.468	

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

It can be seen from Table 3 that the factors that describe the author's rights protection awareness can be summarized into five: induction factor, environmental factor, feedback factor, cost factor, and quality factor. And in Table 4 can get the factor load and the corresponding weight. We calculate the comprehensive score of each factor according to the weight, and use the XGBoost algorithm in the next section to obtain the importance of each factor.

3.3 Factor importance analysis based on XGBoost

3.3.1 XGBoost principle

The full name of XGBoost[24] is (eXtreme Gradient Boosting) extreme gradient boost, which is often used in some competitions, and its effect is remarkable. It is a tool for massively parallel boosted tree, it is currently the fastest and best open source boosted tree toolkit. The algorithm applied by XGBoost is an improvement of GBDT (gradient boosting decision tree), which can be used for both classification and regression problems.

The basic learner in XGBoost can be either CART (gbtree) or linear classifier (gblinear). The steps of the XGBoost algorithm are basically the same as those of GB. They are first initialized to a constant. GB is based on the first derivative ri, XGBoost is based on the first derivative gi and the second derivative hi, iteratively generates a base learner, and adds the update learner. Compared with the traditional learning algorithm, XGBoost considers the case where the training data is sparse, and can specify the default direction of the branch for the missing value or the specified value, which can greatly improve the efficiency of the algorithm. XGBoost also borrows from the practice of random forest and supports column sampling, which can not only reduce overfitting but also reduce calculations, which is also a characteristic of XGBoost that is different from traditional GBDT.

3.3.2 XGBoost Result analysis

This module calls the xgb.feature_importances_ function to find out the importance of each special type. We use the Python language to write algorithm programs to obtain the correlation between each factor and the overall rights consciousness (it should be noted here: the superposition of the factor gain, to determine the gain of a factor, we get the gain corresponding to each tree separately, Then average), we sort it to get Table 5.

Table 5 Ranking of the importance of each factor

index	Inducible factor	Cost factor	Quality factor	Environmental Factors	Feedback factor
Importance	0.2664	0.2261	0.2031	0.1827	0.1218
Sort	1	2	3	4	5

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

It can be seen from Table 5 that the induction factor, cost factor, quality factor, and environmental factor are all important influencing factors in the author's willingness to defend rights, and the feedback factor will play a role in boosting the author's rights protection. First of all, among the inducing factors, each factor is the most direct and easiest factor to be considered by the author in the face of plagiarism, and is the most important factor that can inspire the author's rights protection psychologically. However, such factors include self-induced (the cost of producing original videos) and induced by others (the use of plagiarized videos, the traffic obtained from plagiarized videos). For the author, the intensification of others may increase the author's willingness to defend rights, but for the platform and the UGC environment, this is obviously a bad signal. And self-induction will be an important factor affecting authors' rights protection after plagiarism, but at the same time, plagiarism will input danger signals to authors, which will affect the quality of future original authors. Therefore, to improve the author's awareness of rights protection, you can use the platform to take appropriate incentives to try to ensure self-inducing factors, but we cannot rely too much on inducing factors to try to solve the problem from this aspect. In the cost factor, in the prioritization of the importance of random forests, we learned that the energy cost of protecting rights is a problem that most media creators worry about. Due to the short creation period of such creators, the content of the works is short-term, and the energy is limited. If you fight against the plagiarism incident, it is likely to affect their creation in a long period of time, resulting in greater losses. In terms of the quality factor, the accumulation of copyright knowledge and legal knowledge will change the individual's cognition, making it more sensitive to such events. In terms of environmental factors, if you can get some help in the process of rights protection, it may increase the author's own confidence, and it will also become an important part of the creator's consideration. Therefore, we recommend that the platform proceed from cost factors, quality factors, environmental factors, etc., take these three aspects as the main, and take inducing factors and feedback factors as supplements to take measures.

3.4 Analysis on the construction of the pls-pm model of rights protection willingness factor

In the last part, we divided the 14 indicators that affect the author's rights protection consciousness into five factors including induction factor, environmental factor, feedback factor, cost factor, and quality factor through factor analysis. We found that there is a correlation between the five factors, so we set up Internal matrix, use the R software to draw the preliminary path map of the model (as shown in Figure 2), and constantly amend to establish the author's model of influencing factors of rights protection awareness.

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

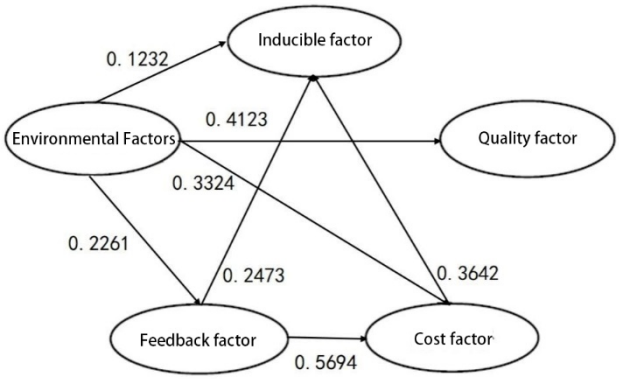

Figure 2 Path map of author's willingness to protect rights

After building the path map, we test the reliability and fit of the pls-pm model, see Table 6 and Table 7 for details.

Table 6 pls-pm model reliability test table

project	Mode	MVs	C.alpha	DG.rho	eig.1st	eig.2nd
Inducible factor	A	3	0.814	0.934	2.146	0.124
Environmental Factors	A	2	0.768	0.907	2.315	0.354
Feedback factor	A	3	0.879	0.916	1.687	0.674
Cost factor	A	4	0.867	0.876	2.334	0.587
Quality factor	A	2	0.845	0.884	2.034	0.635

Table 7 Pls-PM model fit test

project	Type	R^2	Block_Communality	Mean_Redundancy	AVE
Inducible factor	Exogenous	0.000	0.824	0.000	0.824
Environmental Factors	Endogenous	0.134	0.617	0.115	0.743
Feedback factor	Endogenous	0.313	0.653	0.231	0.665
Cost factor	Endogenous	0.446	0.821	0.331	0.798
Quality factor	Endogenous	0.421	0.534	0.247	0.632

It can be seen from Tables 6 and 7 that the C.alpha coefficients of the five factors

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

are all greater than 0.7, the DG.rho values are all greater than 0.8, the first eigenvalues are greater than 1 and the second eigenvalues are less than 1 , Indicating that the model reliability test results are better. The absolute values of the factor loads are all greater than 0.5, and the AVEs are greater than 0.5, which indicates that the scale designed in this paper has good restraint validity. Most of the fitting effect R^2 is greater than 0, indicating that the model has a certain explanatory ability. There may be a certain relationship between the unsatisfactory results and our sample data processing, but it is basically acceptable to consider the overall prediction relationship of the model.

After establishing the path map, the pls-pm model effect is calculated, and the results are shown in Table 8.

Table 8 pls-pm model effect table

variable			Direct effect	Indirect effect	Total effect
Environmental Factors	===>	Inducible factor	0.1232	0.1032	0.2264
Environmental Factors	===>	Feedback factor	0.2261	0.0000	0.2261
Environmental Factors	===>	Cost factor	0.3324	0.1293	0.4617
Environmental Factors	===>	Quality factor	0.4123	0.0000	0.4123
Feedback factor	===>	Inducible factor	0.2473	0.2074	0.4547
Cost factor	===>	Inducible factor	0.3642	0.0000	0.3642
Feedback factor	===>	Cost factor	0.5694	0.0000	0.5694

It can be seen from Table 8 that the continuous development of the self-media creation ecology will have a direct or indirect impact on other factors. Strengthening the joint support of the platform and the audience is the author's all-round demand for rights protection. Earlier we discussed that the induction factor is very important to the author's awareness of rights protection, but the corresponding strategy is difficult to implement, so we hope that the platform will focus on other aspects of improvement. Here we found that the feedback factor and cost factor have a direct impact on the inducing factor, which also confirms our idea of focusing on other aspects. The cost factor is the most important factor other than the inducing factor, and it also needs our special attention. Here, the feedback factor will have a greater direct impact on the cost factor, which also provides a new idea for the platform, that is, to strengthen the feedback experience after the author's rights protection, relatively reducing the actual cost of the author.

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

4. Conclusion and suggestion

4.1 In conclusion

Through the analysis of the differences in the categories of videos produced by the authors, we found that the anime zone and learning zone have the highest rights to protect rights among all video categories. The willingness to defend rights of the Guihu District and the Fashion District is the lowest among all video categories. Through the random forest model, we found that the traffic obtained from plagiarized videos, the production cost of original videos, and the energy cost of defending rights are the most important factors affecting authors' rights protection. After further considering all index information and reducing the dimensionality of all indexes through factor analysis, we combined XGBoost algorithm to find that the induction factor, cost factor, quality factor and environmental factor can directly affect the author's willingness to defend rights. However, for the inducing factor with the highest score, we demonstrated the low implementability of this part, and turned to cost factor, quality factor and environmental factor as the research focus, and the inducing factor and feedback factor jointly promote the improvement of original rights protection awareness. Finally, we use the pls-pm model to find that environmental factors will have a direct or indirect effect on all other factors, feedback factors and cost factors will affect the induction factor, and the feedback factor will also have a direct effect on the cost factor. The platform support provides a broader idea.

4.2 Recommendations based on platform perspective

Since the author belongs to the platform, the biggest change can be made from the perspective of the platform, and the corresponding improvement is made for the author, as follows: 1) According to the second part of the article, the most important for the original author (except for the induction factor, the direction of the induction factor It is difficult to take corresponding measures) The high cost of rights protection, we give corresponding suggestions from the perspective of the difficulty of proof and the large span of rights protection between platforms and authoritative organizations. In view of the difficulty of proof, the platform should strengthen the technical support of electronic proof methods, such as timestamp, blockchain, AI, etc., which may become future proof of ownership and evidence of infringement submitted to the court. In view of the large gap in the rights protection between platforms and authoritative organizations, on the one hand, a unified video copyright registration system, a rapid authorization system, and an original video arbitration system can be established through the establishment of a unified copyright management organization between the platforms to specialize in arbitration of video copyright cases To improve processing efficiency; on the other hand, Internet courts have gradually evolved into a new trend, strengthening the link channels with Internet courts (online courts), by letting authors learn more about such more suitable complaint methods, reducing The cost of defending rights in offline courts

ISSN 2616-5902 Vol. 2, Issue 4: 53-63, DOI: 10.25236/AJBM.2020.020406

reduces the cost of defending original authors' rights and reduces obstacles to defending rights. 2) For the analysis of the research results, we know that the higher the original author's input cost to the work, the stronger the willingness to defend rights. Therefore, in order to improve the original authors' willingness to protect their rights, the platform can strengthen the self-inducing factor, and encourage original creators to improve the quantity and quality of innovative works through original incentives, indirectly increase the production cost of original works, and increase the authors' willingness to protect their rights.

References

[1] Zhu Chen. Research on the Copyright Protection of Online Short Video [J]. Legal System and Society, 2020 (05): 62-63.
[2] Cong Lixian. The core issues of short video copyright protection [J]. Publication Reference, 2019 (03): 1.
[3] Wu Fan, Li Chunzhong, Lin Lifang, Zhu Jiaming. Research on a crowd evacuation simulation algorithm based on cellular automata [J]. Journal of Yanbian University (Natural Science Edition), 2019, 45(04):329-334.