Analysis of Factors Influencing the Effectiveness of Online Video Communication in the Era of Big Data -- an empirical study based on bilibili

Xinyu Ma

Email: mxy1873553751@163.com

University of Electronic Science and Technology of China, Chengdu, 611731

Abstract. With the continuous development of Internet technology and the online video industry has risen rapidly, and watching online videos has gradually become a part of people's daily life. To improve the viewing experience of users, it is necessary to understand which factors can attract users to watch online videos and affect the dissemination effect of videos. Taking the representative video website bilibili as an example, this paper studies the influence factors of online video communication effect from the three dimensions of video content creator, video characteristics and user characteristics, and constructs a model of video communication effect by using multiple linear regression. According to 3804 pieces of data of bilibili platform, the empirical analysis method is used to verify the results. The results show that the video duration has a negative impact on the online video communication effect, while the number of video creator fans, user satisfaction and user participation have a significant positive influence on the video communication effect. The amount of attention of video creators will play a significant negative role in the impact of user satisfaction and the number of fans on the effect of online video communication and play a positive role in the impact of user participation on the effect of online video communication; and different types, different release time of online video communication effects are significantly different. The results expand the re-search in the field of video communication's effect and pro-vide suggestions for individuals and video platforms to improve the effect of video communication.

Keywords: communication effect; user characteristics, data analysis, influencing factors

1 Introduction

With the continuous development of Internet technology, both the application of digital technology and the influence of its communication have undergone tremendous changes in the past ten years. The era of big data takes the "cloud" as the carrier and is characterized by data accumulation, reflecting changes anytime and anywhere. The advent of the era of big data shows that the Internet market is using its database to tap more opportunities, and at the same time the social media industry is developing rapidly, and the development of the video industry is one of the main sources of income in many countries. In the era of big data, online video transmission has shown explosive growth, and users are more willing to focus on products that are closely related to them. In this case, how to collect and integrate data, and apply the results of data analysis to business activities reasonably have become the primary problem that all walks of life need to solve. In recent years, China's online video industry has stepped into the fast lane.

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334320

Both the rapid development of the industry and the growth of the number of users have injected unlimited vitality into the industry. Among them, the younger generation is more willing to share and receive information in the video community. Bilibili is a huge video sharing platform that allows users to post, watch, comment, share and upload videos in real time. It is one of the largest and most popular video websites in China [1]. It has become a multicultural community platform covering more than 7000 circles of the younger generation in China. It has more than 500 million monthly active users and 800 million video views. Therefore, in the process of planning and producing the content of bilibili, through the collection and analysis of the data on the website platform, we can obtain the browsing information and characteristics of users, so as to understand the needs of the audience and carry out accurate network marketing.

2 Literature Review

There has been a lot of theoretical research on the use of video services in big data's era. For example, based on the theory of use and satisfaction, Katz et al explain which service factors the video platform values through the motivation and satisfaction of online video services [2], and even study the motivation of users to use the platform, which limits the specific explanation of how and why users use content [3]. Zhi Zheng et al. analyzed Bilibili’s business model using the business canvas theoretical tool and compared bilibili with other video media such as iqiyi [4]. Based on the technology acceptance model, Davis et al. can not only focus on the perceived ease of use and usefulness of video services in the era of big data, but also reveal users' intention and persistence [5]. Based on the technology acceptance model, Davis et al. can not only focus on the perceived ease of use and usefulness of video services in the era of big data, but also reveal users' intention and persistence[6].

In addition, the researchers also analyzed the usage behavior of video services under the background of big data era from other angles. On the one hand, Li believes that the research focuses more on the use of the platform and the services it provides, as well as perception and satisfaction in the process, rather than content [7]. On the other hand, although some studies on online video focus on content, they tend to focus on specific types of content, such as sport and beauty, or media and sport[8]. It is not enough to fully understand much of the content published in real time, and it is not possible to determine which factors are considered important when selecting and using video content. And how these factors affect the content.

Therefore, for bilibili, which has dominated the form of online video platforms in recent years, in addition to focusing on the factors based on the characteristics of video website platforms under the background of the big data era, this paper will further expand the research on video types. From different perspectives, data analysis is used to study the factors that affect the effect of network video transmission. This research not only has multiple impacts on the bilibili platform and video creators, but also has an impact on the development of various video platforms and related industries.

3 Research Hypothesis

By combing the relevant literature, it can be found that the existing research on the effect of network video transmission is less considering the characteristics of video creators and video types. Combined with the research situation of this paper, the factors that affect the effect of network video communication are divided into three categories: video features, creator characteristics and audience participation characteristics. Following the research of Richier [9], we continue to use the number of viewing time as the evaluation index of video transmission effect.

3.1 The influence of video features on the effect of network video's communication

Video features mainly include video ID, tag, video type, video duration, etc. Compared with traditional text and pictures, online video has rich features, which are the direct factors that affect the audience to watch the video. This article will discuss the effect of video communication in terms of video duration, video release time, and video type. As the main feature of video content, video duration can reflect the content of online videos to a certain extent. As a factor affecting the effect of video transmission, the duration of online video has always been concerned by some researchers. Judging from the release time of the video, 7:00 to 9:00 in a day is the time when most people commute to and from get off work. Many people turn on their mobile phones to watch bilibili online while on the subway. From 11:00 to 14:00, many people's lunch time, most people watch videos again. From 6 pm to 8 pm is the time when many people get off work, and everyone will watch videos. Likewise, from 9pm to 11pm, people are also on their phones before going to bed. In the rest of the time, most people are busy with work or study and have no time to watch online videos, so the number of videos played in the rest of the time is relatively small. As far as bilibili is concerned, some popular activities or events will lead to a surge in the number of certain types of videos, and popular videos will introduce a trend of imitation, and such videos will be watched by viewers. Based on the above analysis, the following hypotheses are put forward:

H1a: The longer the video duration, the worse the video communication effect.

H1b: There are significant differences in video's communication effects with different release time.

H1c: There are significant differences in video's communication effects with different video types.

3.2 The impact of audience participation characteristics on the effect of online video communication

Viewers will preliminaries screen the videos they watch to improve the efficiency of watching videos and meet their own viewing need. Then the interaction between the user and the video is an important reference. Many video platforms do not directly display this project, so the characteristics of user participation are hidden in the number of likes, reports, favorites, coins, and shares of the video. The more satisfied the user is with the video, the better these indicators are, so the information contained in these indicators is called user satisfaction. Bullet screen, a kind of instant comments posted by viewers on the video screen, can not only express the user's own feelings about the video, but also realize the interaction between users; opinions can be considered as the way for users to interact with other users. It not only reflects the user's emotional attitude towards the video, but also indirectly participates in the dissemination of the video. The information is contained in bullet charts and comments are called user engagement. Although high page views may not lead to high positive reviews[10], the converse is not necessarily true.

Moreover, emotional value will indirectly affect the decision of viewing behavior, and the indirect participation of users can also reflect the emotional value that users themselves want to express from the side. Based on the above analysis, the following hypotheses are put forward:

H2a: The higher the satisfaction of the audience, the better the communication of the video.

H2b: The higher the engagement of the audience, the better the communication of the video.

3.3 The influence of the characteristics of video creators on the effect of online video dissemination

In addition to basic information such as name, gender, and profile, the characteristics of the video creator also have some personal dynamics. In 2022, bilibili will update the upper limit of the number of followers. Its purpose is to make it easier for people to pay attention to more and more high-quality up groups on the platform, and further expand the user's own social circle, which means that the more a user follows, the bigger his social circle is, and vice versa. This article uses attention to video creators to measure the size of a publisher's social circle. If the video creator has a large social circle, then regardless of factors such as the video characteristics of the video publisher, the dissemination effect of the video he publishes will be better. When the number of video up-loaders is large, the influence of video features on video dissemination will be weakened [11]. However, when video publishers pay less attention to video websites, users will use information such as video features, number of the likes, and forwarding numbers as signal values to measure video quality, and then decide whether to watch the video. At this time, factors such as the characteristics of the video, the number of likes, favorites, and screen comments will increase the influence of video communication. In addition, Yoganarasimhan et al. used YouTube data to find that the greater the influence of video creators, the better the dissemination effect of online videos, and the dissemination effect of videos released by influential video creators is better than that of users with less influence [12]. The higher the number of fans a video creator has, the greater its influence. At the same time, the number of fans can reflect the breadth of video dissemination from the side. Based on the above analysis, the following hypotheses are put forward:

H3a: The number of followers of the video up-loader plays a negative moderating role in the influence of video features on the effect of video communication.

H3b: The number of followers of the video up-loader plays a positive moderating role in the influence of user characteristics on the effect of video communication.

H3c: the more fans the video creator has, the better the video will spread.

4 Data and variables

4.1 Data Sources

This paper chooses bilibili as the platform of data source. The way of crawling python network data is adopted when obtaining data, and 3900 pieces of data are obtained on February 1, 2023. Excluding null values and repeated values, 3804 valid samples containing 12 features are finally obtained. The descriptive statistical results of the data are shown in Table 1.

Table 1. Descriptive statistics of variables.

Variables	Minimum	Maximum	Mean	Sd
Views	1	25304285.00	270678.74	960666.84
Barrage	0	219270.00	823.78	6057.59
Comment	0	20194.00	379.73	1089.85
Collection	0	862990.00	5184.42	23141.03
Coin	0	2347499.00	4544.96	47273.64
Share	0	200509.00	1312.38	7236.29
Like	0	3027531.00	16313.90	80124.82
Video type	1	13.00	7.44	3.68
Release period	1	5.00	2.46	1.40
Video duration	6	764602.00	1800.38	18578.35

4.2 Variable definitions

First, this paper takes the amount of video playback as the video communication effect index, that is, the dependent variable. Secondly, the factor analysis of the processed data is carried out by using the factor analysis function of SPSS software. As showed in Table 2, the results of KMO test and Bartlett test show that the value of KMO is 0.846, which meets the requirements of factor analysis. At the same time, the results of Bartlett test showed that there was significant difference, which indicated that the data were suitable for factor analysis.

Table 2. Results of KMO test and Bartlett test.

KMO test and Bartlett test		
KMO value		0.846
	Approximately Chi Sqr.	15505.198
Bartlett sphericity test	df	28
	P-value	0.000***

[a] ***represent the significance levels of 1% respectively

In this paper, the principal component analysis is used to extract the factor, and the maximum variance method is used for matrix rotation. Table 3 shows that among the four extracted factors, the first factor has a higher load in the number of collections, coins, shares and likes; the second factor has a larger load in the number of barrage and comment numbers, and the results Matched expectations. Therefore, according to the results of factor analysis and the actual significance of the variables, the score variables of factor 1 and factor 2 are named user satisfaction and user participation, and factors 3 and 4 were still named as video duration and the number of fans. According to the component score coefficient matrix, the expressions of the four factors are given, as showed in equation (1) to equation (4).

$$f_1 = -0.11776667fans + 0.02706418Video\_duration - 0.21694441bullet\_screen \\ -0.13269096comment + 0.33659739collection + 0.36015515coin \\ +0.25555666sharing + 0.31756835like \quad (1)$$

$$f_2 = -0.13459792fans - 0.08479798Video\_duration + 0.7727077bullet\_screen \\ +0.63007784comment - 0.11275183collection - 0.26347033coin \\ +0.08517545sharing - 0.1316689like \quad (2)$$

$$f_3 = 1.05824566fans + 0.019515897Video\_duration - 0.13942315bullet\_screen \\ -0.02478564comment - 0.08471195collection + 0.06561019coin \\ -0.21230472sharing - 0.0296655like \quad (3)$$

$$f_4 = 0.02033932fans + 0.97795215Video\_duration + 0.06743143bullet\_screen \\ -0.19551123comment + 0.04925827collection + 0.07486593coin \\ -0.0563677sharing - 0.00401548like \quad (4)$$

Among them, f_1, f_2, f_3, and f_4 respectively represent user satisfaction, user participation, number of fans, and video duration.

Table 3. The rotated component matrix a.

Variables		Component			
		1	2	3	4
User satisfaction	Collection	0.909	0.246	0.105	0.200
	Coin	0.893	0.132	0.211	-0.092
	Share	0.812	0.352	0.095	0.033
	Like	0.902	0.242	0.191	0.033
User participation	Barrage	0.218	0.822	0.007	-0.042
	Comment	0.388	0.734	0.192	-0.025
Video duration	Video duration	-0.005	0.085	0.003	0.984
Number of fans	Number of fans	0.193	0.199	0.956	-0.001

According to the analysis of the video release time in the research hypothesis, the video release time is taken as a category variable. Therefore, in this paper, the 24 hours of the day are divided into five stages: 7:00 to 9:00 as the first stage, recorded as morning; 12:00 to 14:00 as the second stage, recorded as noon; 18:00 to 20:00 as the third stage, recorded as afternoon; 21:00 to 23:00 as the fourth stage, recorded as evening; and the rest of the 24 hours of the day as the fifth stage, recorded as other times. The corresponding codes of the 5 time periods are respectively recorded as 1 to 5. For video types, according to the collected data, it is automatically divided into 13 categories, namely auto-tune remix, dance, entertainment, technology, food, fashion, automobile, life, film and television, music, games, knowledge, and sports, corresponding to codes 1 to 13 respectively.

5 Results

5.1 Correlation analysis

Figure 1 shows the correlation before the variables are processed, so the relationship between variables can be preliminaries judged. In this paper, Pearson correlation coefficient is used to test the correlation between variables. The results in Figure 1 show that the number of video collections, coins, shares and likes are highly positively correlated with the broadcast volume, but also moderately positively correlated with the number of barrages and comments, and weakly correlated with the video duration. In addition, the correlation analysis of f_1, f_2, f_3 and f_4 is also carried out in this paper. The results show that the correlation between the four variables is less than 0.1, which excludes the interference of multicollinearity and is beneficial to the subsequent correlation analysis.

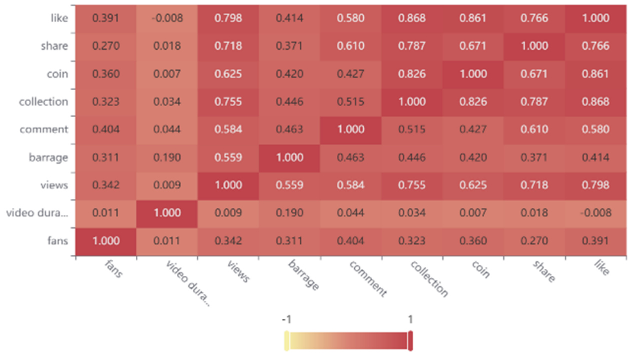

Fig. 1. Heat map of correlation coefficient

5.2 Nonparametric test

The dependent variable in this study is the video playback volume, which is a continuous variable, but for the dependent variable, it is found that it does not conform to the normal distribution and the overall distribution is unknown, so this difference analysis uses the multi-independent sample Kruskal-Wallis test in the non-parametric test method.

According to the fixed category variable (video release time), the quantitative fields (video playback volume) are grouped, and their normality tests are tested respectively to see whether the overall distribution of the data is normal. The results show that based on the number of variable views, the test result P value of grouped variable video release time is 0, less than 0.05. Therefore, the statistical results are important, indicating that there are significant differences in the number of views in different release time periods; however, the difference range Cohen's f value is 0.003, the difference is small. According to the results of multiple comparative analysis after the event, there were significant differences in the number of videos posted between 18:00 and 20:00 and in the evening, noon, and morning, as well as in the rest of the time, but the Cohen's d values were all less than 0.2. Therefore, there is a slight difference in the number of videos released at different times. The result of H1b is valid.

In the same way, this article uses the same method to verify the different analysis of video playback volume of discrete video types. The results in Table 4 show that the Kruskal-Wallis test results show that $P < 0.05$, the result is significant, and there are significant differences in the playback volume of different video types, and the H1c result is established; but the Cohen's f value of the difference is 0.009, which is a very small difference. Further analysis shows that there are significant differences in the playback volume of diverse video types. Compared with game, the video playback volume of ghosts and animals has obvious differences, and the Cohen's d value of the difference is the largest at 0.615, indicating that the difference in video playback volume is 61.5% comes from between diverse groups.

Table 4. Results of Kruskal-Wallis test.

categorical variable	Statistics	P	Cohen's f value
Release period(views)	33.056	0.000	0.003
Video type(views)	83.362	0.000	0.009

5.3 Regression analysis.

After dealing with the variables, combined with the research hypothesis proposed in this paper, a multiple linear regression model affecting the effect of network video communication is established, as showed in equation (5).

$$views = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3 + \beta_4 f_4 + \epsilon \tag{5}$$

Among them, β_0 and ϵ represent the constant term and the error term. Among them, the results of the multiple linear regression model show that model1 $R^2 = 0.684$ in Table 5, indicating that the model has a good fitting effect; The P values corresponding to the coefficients 0, f_1, f_2, f_3, and f_4 are all significant under the test of 0.05, and they all have an impact on the amount of video playback; the coefficients of f_1, f_2, f_3, and f_4 are respectively 650700.066, 441225.961, 113074.757, -19322.256, which means that user satisfaction, user participation and the *number of* fans have a positive impact on video views volume, while video duration negatively affects video playback volume, that is, H1a, H2a, H2b, and H3c are established. The P values corresponding to the coefficients of f_1, f_2, f_3, and f_4 are all significant under the test of 0.05, and they all have an impact on the amount of video playback; the coefficients of f_1, f_2, f_3, and f_4 are respectively 650700.066, 441225.961, 113074.757, -19322.256, which means that user satisfaction, user participation and the number of fans have a positive impact on video views volume, while video duration negatively affects video views volume, that is, H1a, H2a, H2b, and H3c are established.

5.4 Effect analysis

This paper examines the moderating role of the number of followers of video creators in the impact of user satisfaction and user popularity on video playback. The test results are shown in model2 in Table 5. The results in Table 5 show that the R^2 of model2 is 0.09 more than the R^2 of model1, and the P value is less than 0.05. The result is significant. The change in R^2 is meaningful, indicating that the model increases followers user satisfaction degree, followers participation, and followers number of fans have an impact on video views volume. Among them, the coefficients of the three variables of followers user satisfaction, followers participation, and followers number of fans are -403.921, 290.534, and -284.929 respectively, and the P values are all less than 0.05, so these regression coefficients are meaningful, namely The number of video creators' attention has a moderating effect on user satisfaction, user participation, and the number of fans, and the number of followers has a negative regulation on user satisfaction and the number of fans, and a positive regulation on user participation, that is, hypotheses H3a, H3b established.

Table 5. The result of model estimation.

Model	Variable	Unnormalized coefficient β	Standardized coefficient β	T	P	R^2
Model1	(constant)	270678.736	_	27.572	0.000	0.684
	F1	650700.066	0.677	66.272	0.000	
	F2	441225.961	0.459	44.937	0.000	
	F3	113074.757	0.118	11.516	0.000	
	F4	-19322.256	-0.02	-1.968	0.049	
Model2	(constant)	271800.926	_	28.083	0.000	0.693
	F1	740928.315	0.771	45.571	0.000	
	F2	422517.212	0.44	39.738	0.000	
	F3	146778.73	0.153	13.68	0.000	
	F4	-12706.965	-0.013	-1.309	0.191	
	Followers User satisfaction	-403.921	-0.118	-6.806	0.000	
	Followers User engagement	290.534	0.084	6.541	0.000	
	Follow fans	-284.929	-0.097	-7.73	0.000	

6 Conclusion

Taking the representative video website bilibili as an example, based on signal theory, this paper studies the influence factors of online video communication effect from three dimensions: video content creator (the number of up main followers, up main fans), video characteristics (video duration, video release time) and user characteristics (user satisfaction, user participation). And on this basis, a model is constructed by using the effect of multiple linear regression on video communication. The results of empirical analysis show that: first, the length of video has a negative impact on the effect of online video communication, indicating that the communication effect of shorter online video is better than that of longer online video; the number of fans of video creators, user satisfaction and user participation It has a significant positive impact on the video dissemination effect, indicating that as the number of fans of the video creator, user satisfaction and participation in the video increase, the dissemination effect of the online video will increase; secondly, the video creator's The number of followers will play a significant negative regulatory role in the impact of user satisfaction and content creators' attention on the effect of online video dissemination, and play a positive regulatory role in the impact of user participation on the effect of online video dissemination, namely As video content creators have more followers, the effect of user satisfaction and video creators' attention on online video dissemination will weaken, while the role of user participation will increase; thirdly, different types of online videos There are obvious differences in the dissemination effects, and there are also small differences in the dissemination effects of videos released in different time periods; it is further found that among all video types, the dissemination effect of auto-tune remix videos is significantly better than other types of online videos; at 21:00 p.m. Videos released until 23:00 are obviously better than videos released in other time periods. This paper also found that user satisfaction and user participation are the main factors affecting the effect of video communication. Further research found that the type of video will also play a moderating role in the impact of various factors on the effect of video communication.

There are still some limitations in the research of this article: first, this article only studies the data of one platform of the bilibili video website and does not use the data of other platforms. Different video websites may lead to different experimental results; second, this research on the effect of network video dissemination did not consider the time factor and user's emotional factors; follow-up research can take these aspects into account, and the video dissemination effect for a more comprehensive analysis.

References

[1] Ming, Y., Minqi, C., Zejiu, G., Ziyi, Y. and Kunxia, L. (2022) of Conference Research on the Business Model of Bilibili and Its Current Problems. in Proceedings of the 2022 International Conference on Creative Industry and Knowledge Economy (CIKE 2022). QINGDAO. p. 261-266. https://doi.org/ 10.2991/aebmr.k.220404.048.

[2] Katz, E., Blumler, J. G. and Gurevitch, M. (1973) Uses and gratifications research. Public Opinion Quarterly. **37**(4): p. 509-523. https://doi.org/ 10.1086/268109.

[3] Chiang, H.-S. and Hsiao, K.-L. (2015) YouTube stickiness: The needs, personal, and environmental perspective. Internet Research. **25**: p. 85-106. https://doi.org/ 10.1108/IntR-11-2013-0236.

[4] Zhizheng, Z., Li, Z. and Yajun, L. (2020) in AHFE 2020 Virtual Conferences on Usability and User Experience, Human Factors and Assistive Technology, Human Factors and Wearable Technologies, and Virtual Environments and Game Design, 16-20 July 2020. Editor^Editors, The Analysis of Influence Factors of User Experience in Bullet Screen. Springer International Publishing, USA. p. 272-278. https://doi.org/ 10.1007/978-3-030-51828-8_36.

[5] Davis, F. D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. **13**(3): p. 319-340. https://doi.org/ 10.2307/249008.

[6] Cha, J. (2014) Usage of video sharing websites: Drivers and barriers. Telematics and Informatics. **31**: p. 16–26. https://doi.org/ 10.1016/j.tele.2012.01.003.

[7] Li, J. Y., Wang, Z. H., Li, Z. Y. and Ieee (2019) of Conference Understanding the Usage Characteristics of Personalized Online Video Services. in IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Paris, France. p. 786-791. https://doi.org/ 10.1109/INFCOMW.2019.8845313.

[8] Borwankar, V., Chris, C., Kumar, H. and Rahaman, S. (2022) of Conference Sensitivity Analysis using Regression Models: To Determine the Impact of Meta-level Features on the Youtube Views. in Proceedings of the 11th International Conference on Data Science, Technology and Applications. Lisbon, Portugal. p. 270-275. https://doi.org/ 10.5220/0011207300003269.

[9] Richier, C., Altman, E., Elazouzi, R., Jimenez, T., Linares, G. and Portilla, Y. (2014) of Conference Bio-inspired models for characterizing YouTube viewcout. in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014). p. 297-305. https://doi.org/ 10.1109/ASONAM.2014.6921600.

[10] Chang, W. L. (2018) of Conference Will Sentiments in Comments Influence Online Video Popularity? in 2018 IEEE International Conference on Big Data (Big Data). Seattle, WA, USA. p. 3644-3646. https://doi.org/ 10.1109/BigData.2018.8621938.
[11] Seiler, S., Yao, S. and Wang, W. (2017) Does Online Word of Mouth Increase Demand? (And How?) Evidence from a Natural Experiment. Marketing Science. **36**(6): p. 838-861. https://doi.org/ 10.1287/mksc.2017.1045.
[12] Yoganarasimhan, H. (2011) Impact of social network structure on content propagation: A study using YouTube data. Quantitative Marketing and Economics. **10**(1): p. 111-150. https://doi.org/ 10.1007/s11129-011-9105-4.