Research on the Evaluation Model of Mobile Application

Fengyang Liu*, Honglei Xu[a], Zhaoxi Wang[b]

China Mobile Research Institution, Beijing, China
*liufengyang@chinamobile.com, [a]xuhonglei@chinamobile.com,
[b]wangzhaoxi@chinamobile.com

*Corresponding author Email: liufengyang@chinamobile.com

ABSTRACT. With the development of mobile Internet and the peak of user growth, mobile application development has shifted from brutal expansion to refined operation. In order to better adapt to mobile application development and user needs, this paper constructs a comprehensive product development evaluation system model from the perspective of product and operation based on existing research. The model comprehensively considers the influencing factors such as operational capability, competitiveness, product usability, and determines the weight of the indicators by combining qualitative and quantitative methods. On the basis of model construction, empirical research is conducted on video applications, iQiyi, Youku, Tencent video and Migu, to analyze their development level, diagnose problems and put forward suggestions.

Keywords: mobile application, evaluation model, video application, Analytic Hierarchy Process (AHP), CRITIC, entropy method

1 Introduction

With the decline of the mobile internet traffic dividend, the user base and market for various mobile applications(apps) have gradually reached a saturation point. In 2022, the overall scale of mobile internet users in China was 1.065 billion, with a year-on-year growth of only 3.5%. However, the average weekly mobile usage time per user was 26.7 hours, showing a 6% decrease compared to the previous year. Users are shifting between apps rather than extending their overall usage time, which has contained the brute scale development. Therefore, how to achieve better market performance and user reputation through refined operations and optimized user experiences has become a key issue that mobile apps need to address.

In this trend, the evaluation for apps development should also adapt accordingly. Currently, both target-oriented evaluations and market value assessments emphasize apps utility and experience, which is not comprehensive enough. They can easily lead to a situation where apps always focus on quantity over quality or lost in unlimited user demands，which hinders the development of mobile apps to a certain extent. Therefore, this article approaches the problem from a balanced perspective, considerations of market results, operational processes, apps experiences, and reputation. By employing

G. Guan et al. (eds.), *Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023)*, Atlantis Highlights in Computer Sciences 16,
https://doi.org/10.2991/978-94-6463-264-4_19

a combination of qualitative and quantitative methods, a comprehensive and scientific product development evaluation system is constructed.

2 Literature Review

Since the advent of the mobile internet era in 2011, numerous scholars both domestically and internationally have conducted research on the evaluation of mobile apps. They primarily focuses on two perspectives: the product perspective and the operational perspective.

2.1 Research from the Product Perspective

The product perspective refers to the evaluation conducted based on the functionality performance of apps. Scholars began researching product evaluations during the early stages of mobile internet development. Huy and Vanthanh evaluated different forms of mobile apps, such as native apps and HTML5, focusing on three dimensions: usability, functional design, and downloading and updating from the user's perspective [1]. Huang Wei, Li Zongke, and Huang Jianqiao identified user experience and functionality as the most important factors influencing the quality of an app. They constructed a universal apps evaluation model from four dimensions: app downloads and comments from app stores, user experience, visual, and network[2]. With the rapid development of the Chinese internet industry, domestic scholars have proposed more evaluation models for different subcategories of apps. Jiang Yue developed 14 related indicators to evaluate social apps based on usability-related research theories. Through questionnaire surveys and interviews, they found that eight indicators, including product errors, learnability, and satisfaction, had the great impact on evaluation results [3]. Liu Mingzheng and others focused on reading products and, based on the Technology Acceptance Model, found that ease of use was less important than usability factors such as interface aesthetics and operational stability. They emphasized that evaluation should pay more attention to factors representing usability, such as content resource quality, interactive functionality, and reputation [4].

2.2 Evaluation Research from the Operational Perspective

The operational perspective refers to the evaluation conducted based on the effectiveness of specific workflow, such as user operations and content operations. The well recognized AARRR product operation model proposed by Dave McClure in 2007 has been widely applied in various mobile apps. It describes user operation from five aspects: Acquisition, Activation, Retention, Revenue, and Referral, forming a complete analysis method and indicator system. In recent years, domestic scholars have made many attempts to construct the evaluation systems based on the development of mobile app. Zhang Tingquan proposed the POMP evaluation model based on the characteristics and common issues of mobile app. The model includes four dimensions, product capability, operational capability, marketing capability, and market performance, to

evaluate product health [5]. Yang Yang and others describe tourism app operation from five stages, problem recognition, information search, solution evaluation, purchase decision, and post-purchase behavior, base on user purchasing decision-making paths theory. They designed evaluation indicators to evaluate operational effectiveness in each stage comprehensively[6].

By analyzing the above research results, it is found that the app evaluation model involves multiple perspectives and disciplines. Most research results focus on a single field, with fewer studies constructs evaluation model from multiple perspectives. Therefore, this study draws on existing research achievements, considering both the product and operational perspectives, and proposes an innovative evaluation model that includes competitiveness, product usability, and operational capability. Meanwhile, an empirical study is conducted using top video apps as an example to provide guidance for the healthy development of apps.

3 Research Methodology

In order to construct a comprehensive and scientific evaluation model for app development, this study considers the factors from both the product perspective and the operational perspective. It establishes product usability evaluation indicators based on product functionality and performance experience, and operational capability evaluation indicators based on the AARRR model. Additionally, two indicators, user market share and reputation evaluation, are added as competitiveness evaluation indicators. Finally, this study selects top video apps for empirical research, using a combination of subjective and objective methods. The Delphi method and Analytic Hierarchy Process (AHP) are initially used to determine the weights of the dimensions, and then the CRITIC method and entropy method are employed to determine the final weights of the indicators based on data volatility and independence.

3.1 Empirical Research Object selection

Considering the effectiveness of the model and credibility and availability of the data, the target apps for empirical research should have a large user base, a high market maturity, and readily available data. Therefore, this study selects four mobile video apps, namely iQiyi, Youku, Tencent Video, and Migu Video, from the app store.

3.2 Competitiveness Indicators

Competitiveness mainly reflect the final results of market operations and include two indicators: user market share and reputation. User market share is measured by the market share of new users and the market share of cumulative users. Reputation is calculated by combining indicators such as net positive review rate and sentiment score from app store reviews.

3.3 Product usability Indicators

Product usability include product functionality score and performance experience score, both obtained through product inspection and experiential evaluations. The product functionality score is determined based on dimensions such as effectiveness, completeness, and error rate, while the performance experience score is determined by speed, latency, and power consumption.

3.4 Operational Capability Indicators

Operational capability is mainly built around the AARRR framework. Acquisition is measured by the monthly number of new users, activation is measured by the monthly number of active users and monthly average usage duration, and retention is measured by the retention rate of the next month. Revenue and Referral are not included in this study due to their high correlation with other indicators. Firstly, the price levels of video apps are generally consistent, so revenue is highly correlated with user scale. Secondly, user referral is conceptually similar to reputation. The data for operational capability indicators are derived from DPI analysis of user mobile internet behavior.
In conclusion, the indicator system of the app development evaluation model constructed in this study is shown in Table 1:

Table 1. Indicator System

Primary Dimension	**Secondary Dimension**	**Indicator Name**
Operational Capability	Acquisition	Number of New Users
	Activation	Activation Rate
		Monthly Average Usage Duration
	Retention	Retention Rate of the Next Month
Competitiveness	Market Share	Market Share of New Users
		Market Share of Scale
	Reputation	Reputation Score
Product usability	Functionality	Functionality Score
	User Experience	User Experience Score

4 Modeling

4.1 Determining Dimension Weights based on the Analytic Hierarchy Process (AHP)

To quantitatively explore the impact of each dimension on apps development, this study first utilized the Delphi method to obtain expert ratings. The expert group consisted of product managers, executives, independent researchers, and regular users in

the field of video apps, all of whom had several years of experience in the industry or using video apps, ensuring the validity, accuracy, and comprehensiveness of the data.

The process of determining dimension weights based on the AHP consists of three main steps.

Step 1: Constructing a judgment matrix using the Delphi method. Experts are required to compare the importance of each indicator dimension pairwise using a 10-point scale. The average scores from all experts are used to construct the judgment matrix.

Step 2: Calculating the eigenvector and dimension weights. The calculation results are shown in Table 2:

Table 2. Dimension Weights

Primary Dimension	**Secondary Dimension**	Eigenvector	Weights (%)
Operational Capability	Acquisition	1.621	21
	Activation	1.688	6
	Retention	1.279	6
Competitiveness	Market Share	1.607	8
	Reputation	0.529	22
Product usability	Functionality	0.52	21
	User Experience	0.646	16

Step 3: Consistency check analysis. Due to the large number of indicator dimensions, in order to prevent experts from introducing arbitrariness and to ensure logical consistency in the scoring process, a consistency check is performed on the judgment matrix. This study adopts the CR (Consistency Ratio) method for the check, with the following formula:

$$CR = \frac{CI}{RI} \tag{1}$$

The CI (Consistency Index) reflects the degree of consistency, and a smaller CI value indicates better consistency. RI (Random Index) is a fixed value obtained from a reference table, representing the average random consistency. Typically, if CR < 0.1, the consistency check is considered satisfactory.

After calculation, the result is shown in Table 3. The CR value in this study is less than 0.1, meeting the requirements and passing the consistency check.

Table 3. Consistency Check Result

CI	RI	CR	Result
0.063	1.341	0.047	PASS

4.2 Determination of Indicator Weights based on CRITIC Method and Entropy Weight Method

Since there are two indicators under the dimensions of market share and activation, it is necessary to allocate weights to each indicator reasonably. Considering that indicators with greater dispersion and volatility have a greater impact on the comprehensive evaluation results, while indicators with higher correlation have a smaller impact, the CRITIC method and entropy method are used to comprehensively determine the specific indicator weights. The entropy weight method can measure the dispersion of indicators, while the CRITIC method can measure the volatility and the correlation of indicators.

Step 1: The entropy method is applied to calculate the each indicator weight. The results are shown in Table 4:

Table 4. Indicator Weights by Entropy Method

Indicator	Entropy	Weights(%)
Activation Rate	0.88	24.695
Monthly Average Usage Duration	0.635	75.305
Market Share of New Users	0.89	33.567
Market Share of Scale	0.783	66.433

Step 2: The indicator weights were calculated using the CRITIC method. The results are shown in Table 5:

Table 5. Indicator Weights by CRITIC Method

Indicator	Weights (%)
Activation Rate	44.694
Monthly Average Usage Duration	55.306
Market Share of New Users	48.8
Market Share of Scale	51.2

Step 3: The final weights of the indicators were obtained by averaging the results of the two methods and combine with the subjective dimension weights. The results are shown in Table 6:

Table 6. Weights of all Indicators

Primary Dimension	**Secondary Dimension**	**Indicator Name**	**Weights** (%)
Operational Capability	Acquisition	Number of New Users	21.54
	Activation	Activation Rate	7.29
		Monthly Average Usage Duration	13.73
	Retention	Retention Rate of the Next Month	16.18

Competitiveness	Market Share	Market Share of New Users	8.43
		Market Share of Scale	12.04
	Reputation	Reputation Score	6.45
Product usability	Functionality	Functionality Score	6.38
	User Experience	User Experience Score	7.96

In the evaluation model of video app, the importance of the primary dimension is ranked as follows: operational capability > competitive capability > product capability. From operational capability perspective, the weights of indicators such as the number of new users, monthly average usage duration, and the retention rate in the next month are relatively high. This indicates that ensuring user engagement and sustained product usage are crucial aspects of product development. From competitiveness, the market share of scale is the most important, as it reflects the overall strength of the product. In the product usability, since video apps are relatively mature and have a high degree of functional homogeneity, performance factors such as speed and latency exhibit more distinct differences.

5 Empirical Research - Analysis of Operational Development Level for Top Video Apps

Based on the model constructed in this study and considering the data and evaluation indicators, the assessment results for the top video apps are calculated and presented in Table 7:

Table 7. Empirical Result of Video Apps

Empirical Result		Tencent Video	iQiyi	Youku	Migu
Operational Capability	Acquisition	43.8	50	45.7	55.9
	Activation	15.1	16.1	14.0	12.8
	Retention	62.3	63.8	62.7	63
	sub-score	34.6	37.2	34	35
Competitiveness	Market Share	17.5	14.1	15.5	13.0
	Reputation	99	80.4	91.9	92.8
	sub-score	23.9	19.3	21.4	18.9
Product usability	Functionality	89.8	93.1	91	91.5
	User Experience	78.1	90.5	89.9	82.9
	sub-score	11.9	13.1	13	12.4
Overall		70.5	69.6	68.4	66.4
Rank		1	2	3	4

Tencent Video performed the best among video apps. Specifically, it excelled in competitiveness, far surpassing other products, especially in terms of product reputation score, indicating its ability to meet user demands and generate positive recom-

mendations. Its strong reputation also helped Tencent Video gain a leading position in the market. iQiyi demonstrated outstanding operational capability, ranking first in terms of monthly average usage duration and the retention rate in the next month. However, iQiyi did not receive user recognition in terms of product reputation, and the lack of user recommendations hindered its ability to expand its product influence, finally affecting its market share. Youku had an advantage in product usability but was limited by poor performance in operation such as acquisition, activation, and retention. It failed to fully leverage its product strengths and needs to focus on optimizing its operational processes and efficiency. Migu leveraging the advantages of China Mobile's user resources, maintained a leading position in acquisition and performed well in reputation through the operation of a series of popular sports content such as the World Cup and Winter Olympics. However, there is still a gap compared to other apps' market share. This is mainly due to excessive dependence on periodic sports events, leading to a high churn rate during non-event periods, making it difficult to achieve long-term and stable market share growth.

6 Conclusion

This paper complements the lack of previous research of app evaluation model, considering the characteristic of both mobile apps and operation works. Based on constructing a comprehensive evaluation model, top video apps are compared and evaluated. Through a combination of qualitative and quantitative methods, the strengths and weaknesses of different video apps are fully explored. This can effectively help companies understand user preferences, clarify market positioning, enhance product quality, and contribute to the healthy and sustainable development of the video apps market.

7 References

1. Huy N P, Vanthanh D. Evaluation of mobile app paradigms[C]// International Conference on Advances in Mobile Computing & Multimedia. 2012.
2. Huang W, Li Z, Huang J. Research on evaluation index system of mobile applications (APP)[J]. Books and Information, 2016(3):8.
3. Jiang Y. Research on the usability evaluation of social mobile apps from a user perspective[D]. Nankai University, 2016.
4. Liu M, Qiu Z. Reader demand and market satisfaction: Research and application of evaluation system for reading apps[J]. Library and Information Science Research, 2020(8):12.
5. Zhang T. Discussion on the evaluation system of mobile internet product operation based on user experience[J]. Telecommunications Science, 2011, 27(8):4.
6. Yang Y, Yang S, Zhang M, et al. Evaluation index system for travel booking apps based on DEMATEL method[J]. Journal of Tourism, 2016.