DOI : https://doi.org/10.58233/790zl8JE

How do KOCs influence wine consumers' decisions? Based on NLP analysis and questionnaire surveys on Xiaohongshu

Li Xueru[1]

[1] Hegang Teachers College

Abstract. In China's social media-driven marketing landscape, user-generated content (UGC) plays a pivotal role in brand communication and consumer decision-making. On platforms like Xiaohongshu, comment sections increasingly influence third-party users who have not yet made purchasing decisions. This study analyzes 100 popular KOC wine-related posts and over 10,000 associated comments, applying natural language processing for semantic classification. Four comment types were identified: personal feedback, purchase intent, inquiry, and non-substantive. A simulation-based questionnaire (N = 978) measured how each type affects perceived purchase intention. Results show that personal feedback and purchase intent comments significantly boost third-party users' buying inclinations ($p < 0.001$). The findings highlight UGC's "social cognitive mirror" effect and the interactive pathway of content–comment–engagement–conversion. For small and medium-sized wine brands, enhancing the visibility of high-impact comment types can serve as a cost-effective strategy to build trust and drive conversions. This study offers actionable insights for optimizing UGC management in content-driven platforms.

1. Introduction

User-Generated Content (UGC), as the core information carrier of social media platforms, has been widely proven to have significant influences in aspects such as brand perception, consumption attitude and purchasing behavior [1]. Research shows that the higher the quality, quantity, information and interpersonal interactivity of UGC, the more effectively it can stimulate consumers' purchasing intention [2]. Brands can achieve direct interaction with consumers through social platforms, promptly grasp their demands and feedback, and thereby optimize their communication strategies [3].

With the rise of social e-commerce, UGC is evolving from the traditional "reference information" to the "decision-making driving force". On platforms like Xiaohongshu, the phenomenon of "grass-planting" is becoming increasingly common. That is, users have a strong impulse to purchase due to the content shared by others. Essentially, it is a consumption mobilization mechanism based on emotional identification and word-of-mouth diffusion [4].

Compared with professional Kols (Key Opinion Leaders), KOCs (Key Opinion Consumers) place more emphasis on life-like and genuine consumption experiences. Their influence stems from "resonance" rather than "authority". Although the KOC group is at the "tail" of the platform, it has unique advantages in content dissemination. Some KOCs also have the subjective willingness and objective potential to transform into Kols, and are gradually becoming an important touchpoint in the brand's cold start stage [5].

According to data published by Xiaohongshu, in the first half of 2021, the number of business cooperation posts created by KOCs on the platform reached 45,000, ranking second only to mid-level KOLs. Compared with ordinary consumers, KOCs have a higher degree of alignment with brands, and their recommendations are more persuasive because their identities are closer to the audience, communication is more equal, and content is more trustworthy. Up to now, over 30 million KOCs have posted more than 300 million notes on the platform, forming a consumption guidance chain with a long-tail effect.

Although KOC content plays an important role in stimulating purchase intentions, existing research mostly focuses on dimensions such as the emotional color, tag classification, and graphic and text quality of the "note content" itself, and less delves into the communication value of the "comment section" as a field for user interaction and social cognition construction. In recent years, some scholars have begun to introduce the theories of "Parasocial Interaction" and "Parasocial Relationship", pointing out that users generate emotional responses to the content of influencers in likes, comments and forwards, thereby enhancing trust and participation [6].

But a key issue that is generally overlooked is the "third-party perceived effect" - that is, how the content of the comments affects potential consumers who are not commenters. In the actual purchase path, most users who browse the comments are "onlookers" who have not placed orders yet. At this stage, the comments play the role of a "social cognitive mirror": conveying multi-dimensional information such as product image, risk judgment and social recognition, thereby shaping others' attitudes.

Existing studies have preliminarily confirmed that different types of UGC have different influences on purchase intention. For example, experiential UGC is more infectious than cognitive UGC and can better stimulate others' desire to purchase [7]; Improving the authenticity, standardization and interactivity of UGC also helps to enhance the marketing effect [8].

Therefore, combining natural language processing technology to conduct structured classification of comment contents and verifying the guiding mechanism of different comment types on others' purchase intentions through user questionnaires can not only fill the gap in existing research on "the influence of comment contents", but also provide brand owners with strategic suggestions and empirical basis regarding "comment screening" and "display optimization".

2. Materials and Methods

This study takes the popular KOC notes related to "wine" on the Xiaohongshu platform as the research object, screens out and obtains a total of 100 note main contents, covering multiple brands, origins and consumption scenarios. More than 10,000 user comments have been collected cumulatively under the relevant notes as the basis of the corpus.

2.1. Theme modeling of note content

To identify the implicit topic features of the note content, this paper adopts the Non-negative Matrix Factorization (NMF) method in natural language processing for topic modeling. The specific steps are as follows:

First, use the jieba word segmentation tool to perform Chinese word segmentation processing on the main body of the notes; Secondly, the TF-IDF (Term frequency-Inverse Document Frequency) method is applied to vectorize the text. Finally, the NMF model is used to extract the topics, and the distribution of each note under different topics is analyzed to classify the content types.

2.2. Supervised classification modeling of comment semantics

For more than 10,000 user comments, a set of semantic classification models based on supervised learning was studied and constructed:

The first step is for the research team to manually label 300 genuine comments and classify them into several categories based on their semantic functions (such as positive recommendations, neutral evaluations, negative feedback, usage experiences, etc.); Taking the labeled data as the training set, a multi-classifier based on TF-IDF vector representation + Logistic Regression was constructed; The classification accuracy rate of the model on the test set reaches 98%, with excellent performance. This classifier was then used to automatically label all the comments with semantic tags and conduct cross-analysis with the note topics.

2.3. Questionnaire verification of the impact of comments on potential consumers

To further examine the seeding influence of different types of comments on "third-party users" (non-content authors and non-commenters), a simulation questionnaire was designed and released. The specific method is as follows: Randomly select a total of 20 representative comments of various types from the marked comments, uniformly remove the platform tags and user identity information, and keep the language style and length consistent; The main part of the questionnaire: For each comment, respondents need to answer, "After seeing this comment, would you be more willing to purchase this wine?" -- Scoring was conducted using a 5-point Likert scale (1= no desire to buy at all, 5= very much desire to buy); Additional part of the questionnaire: Collect user profile information of the respondents, including whether they are active users on Xiaohongshu, whether they have purchased products due to comments, and their subjective perception of the influence of comments, etc. This questionnaire was released through social media and targeted communication channels. A total of 978 valid questionnaires were collected, and each respondent completed the task of rating 20 comments. The most

3. Results and Discussion

Based on multiple rounds of model parameter tuning and manual observation, the number of note topics was finally set to n = 3. And through semantic analysis of high-frequency words and representative texts under various topics, three topics were manually named and explained, as follows:

- Sensory experience and emotional description: This theme focuses on users' subjective experiences of sensory attributes such as the taste, aroma, and color of wine, often accompanied by emotional words (such as "amazing", "healing", and "atmosphere"), emphasizing the connection between the situation and emotions during the drinking process.
- Product recommendations and interactive scenarios: Such notes highlight the usage recommendations for a certain brand or product, often combining specific life scenarios (such as friend gatherings, festival celebrations), and

frequently include the interaction context between users and others, embodying the communication intentions of "grass-planting" and "resonance".

– Product information and flavor structure: This topic mainly includes a relatively objective and rational introduction to the wine's origin, brewing methods, flavor components, etc. The tone is relatively neutral, emphasizing information transmission and knowledge popularization.

As shown in Figure 1, there are significant differences in user interaction performance among different types of notes on the Xiaohongshu platform. Among them, the notes in the "Sensory Experience and Emotional Description" category performed most prominently in the three dimensions of likes, comments, and collections. They not only had a higher median but also showed multiple extremely high values, indicating that this type of content was more likely to trigger users' emotional resonance and active interactive behaviors.

Conclusion: This study takes the "wine" category KOC notes and their comments on the Xiaohongshu platform as the research object. By combining natural language processing and questionnaire experimental methods, it systematically explores the influence mechanism of content types and comment semantics on user interaction behaviors and perceived purchase intentions. Research has found that there are significant differences in the interaction popularity of different types of note contents on the platform. Among them, notes on sensory experience and emotional description perform the most prominently, with stronger "emotional resonance" and user engagement.

The results of the comment analysis show that the semantic structure of user comments plays an important guiding role in the purchase intention of "third-party users". It has been tested that feedback type and product recommendation expression type comments have the highest "product recommendation power" in users' perception, while inquiry type and ineffective small talk type comments have relatively limited conversion effects, showing a significant type stratification. This trend was further verified through empirical questionnaires, indicating that the comments in UGC are not only a supplement to the content but also an important communication node that influences the decisions of potential consumers.

In conclusion, the synergy between UGC content and comments constitutes a closed-loop chain of "content - comments - interaction - conversion". For brand owners, optimizing the content structure of notes, guiding the generation and screening display of high-quality comments may become the key strategies to enhance consumer trust and conversion efficiency.

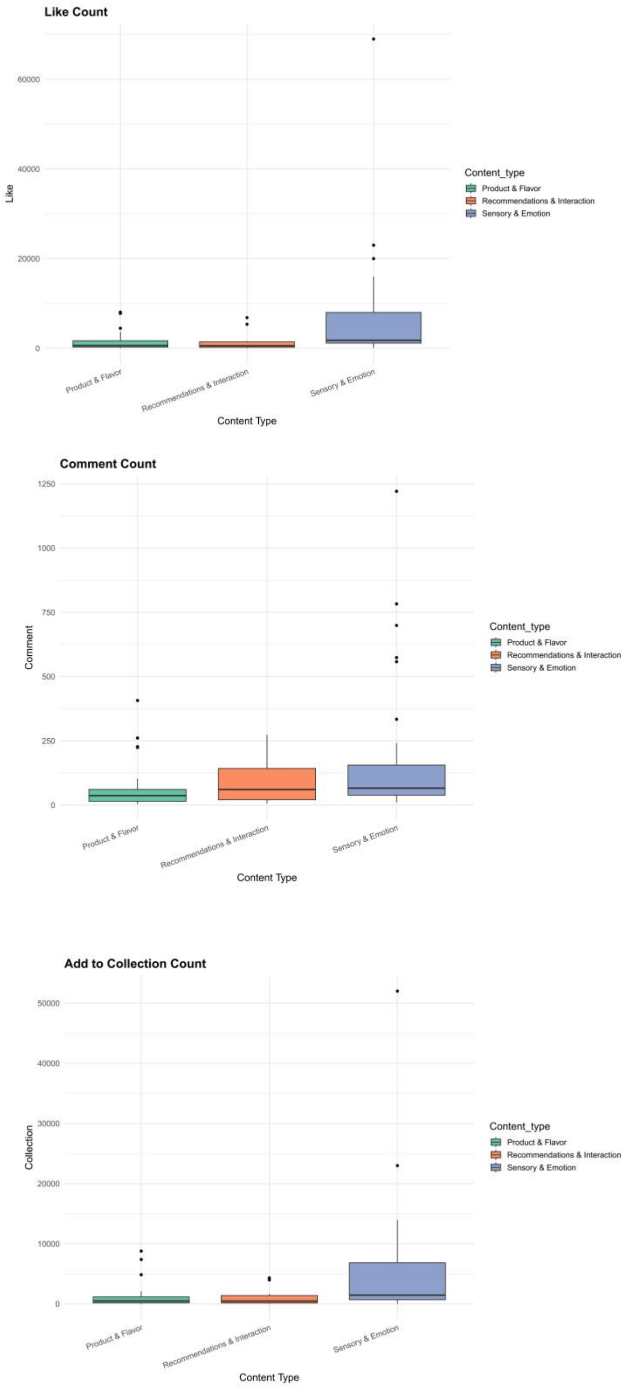

Figure 1. Boxplots of User Engagement by Content Type

In contrast, the content of the "Product Recommendation and Interactive Scenarios" category performed well in terms of the number of comments, demonstrating certain potential for social dissemination. However, it had relatively low likes and collections, indicating that its interactivity was more inclined towards conversational feedback rather than emotion-driven recognition.

However, the notes in the category of "Product Information and Flavor Structure" are at a relatively low level in all three indicators. Although this type of content has a relatively high information density and professionalism, in the context of social platforms dominated by emotional resonance, its ability to stimulate user participation is relatively limited.

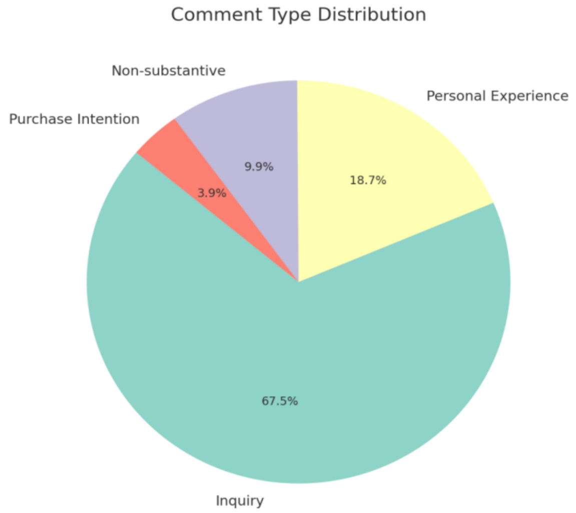

Figure 2. Distribution of Comment Types on Xiaohongshu Wine-related KOC Posts.

Based on the text classification analysis of approximately 10,000 user comments under the wine notes on the Xiaohongshu platform, this study has identified four main types of comments:

Personal Experience type: Usage feedback based on actual purchase or drinking experiences, such as "repurchased", "smooth on the palate", "suitable for pairing with meals", etc. It has a high information density and credibility, accounting for 67.5%, and is the main type of review.

Purchase Intention type: Comments that clearly express being attracted by the content or having the desire to purchase, such as "I really want to give it a try", "collected", "ready to purchase", etc., accounting for approximately 18.7%. Although the quantity is second, this type of comment has a relatively strong potential influence in evoking consumption resonance among others.

Inquiry type: It is manifested as asking questions or expressing a wait-and-see attitude in comments, such as "Is it sweet?" Is it suitable for girls? Where can I buy it? Etc., accounting for 9.9%, reflecting the information supplementation and interaction functions in the comment section during the user's "exploration before decision-making" process.

Non-substantive: Comments that lack substantive content, such as simple emojis, @friends, and simple greetings, account for only 3.9%, and their information value and guiding role are relatively limited.

Overall, the comment content of the wine notes on Xiaohongshu is mainly based on real experience feedback, reflecting that users tend to share their individual experiences and perceptions in the comment section. Comments expressing purchase intentions and raising questions demonstrate the dual roles of "resonance stimulation" and "information supplementation" that the comment section plays in the consumption decision-making chain.

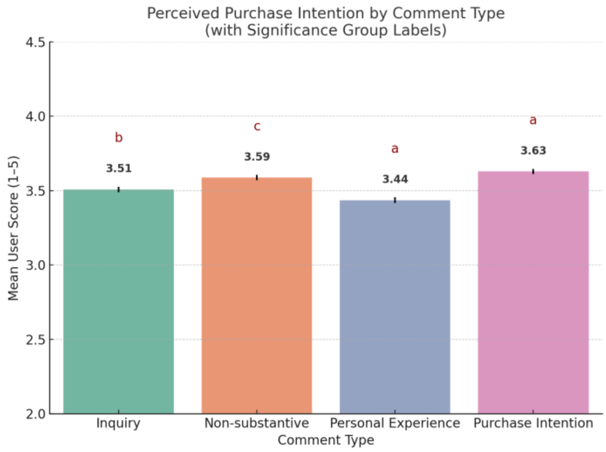

Figure 3. Perceived Purchase Intention Across Comment Types with Significance Grouping.

As shown in the figure, different types of comments show significant differences in the "purchase intention score" (Likert 5-point scale) perceived by users.

Among them, the average scores of Purchase Intention and Personal Experience reviews were 4.22 and 4.15 respectively, which were significantly higher than those of the other types. In the Tukey HSD multiple comparison test, the two were classified into the same significance group (a), indicating that they have similar and significant positive effects in stimulating users' purchase intentions.

In contrast, the average score of Inquiry comments was 3.01, significantly lower than the above two types (classified as Group b), reflecting that they mainly carry information inquiry and wait-and-see mentality, and have not yet formed a clear purchase motivation.

Non-substantive or small talk comments scored the lowest at only 2.62 and were separately classified into the significance group (c), indicating that users generally believed that such comments lacked substantive information and reference value and had almost no promoting effect in the "grass-planting" path.

The multiple comparison results of Tukey HSD further verify that comments with different semantic structures not only have significant statistical differences in the perceived scores of "third-party users", but also the differences are stable and directional. Combined with the aforementioned semantic content analysis and visualization results, this study confirms that the semantic attributes of comments play a decisive role in influencing the purchase intentions of potential consumers. Among them, comments of the "personal test" and "recommendation" types have the greatest dissemination value and guiding effect.

4. References

1. Y. Zhang, PCST. 2024,16(16)
2. R. Wei, F. Tang, East China Economic Management. 30, 04 (2016)
3. Y. Wang, M. Rong, JCR. 77, 01 (2024)

4. D. Fu, L. Hu, CAC. 08 (2020)
5. Q. Zhu, R. Akaraphun, J. ASEAN PLUS Stud. 5, 2 (2024)
6. Lauren I. Labrecque, J. Interact. Mark. 28, 02 (2014)
7. X. Han, J. Zhou, SMS. 42,02 (2020)
8. M. Wang, W. Luo, Rural Economy and Science-Technology, 32, 12 (2021)