

Setting Market Rules to Combat Fake Positive Reviews Based on Game Theory: Achieving a Win-win Situation for Consumers, Government and Merchants

Jiachen Han

International College Beijing, China Agricultural University, Beijing, China

hanjiachen13522812071@gmail.com

Abstract. In the era of online shopping, the phenomenon of false positive reviews for online products is emerging in an endless stream, seriously disrupting the market order. In order to protect the rights and interests of consumers, traditional governance concepts mostly deal with chaos from the perspectives of state fines and supervision and improving consumers' ability to discern. This study innovatively breaks through the binary opposition paradigm of "government regulation-market entities". While improving traditional governance measures, it proposes an anti-false positive review scheme that maximizes reliance on market autonomy rather than government regulation and allows merchants to make profits. This paper first analyzes the pros and cons of traditional governance measures. Then, by analyzing problem examples and references, a four-stage scheme is used to integrate the governance issues of various entities. Finally, the cost differentiation method of the signal game model is designed to study the effectiveness and feasibility of the new scheme. Therefore, taking into account the interests of all parties and seeking a win-win solution can provide a better solution for the governance of false positive reviews.

Keywords: Signal Game, False Positive Reviews, Trust Crisis, Market Rules.

1. Introduction

The problem of fake positive reviews on e-commerce platforms is becoming increasingly complex, showing a diversified development trend from low-disguise fake reviews to high-disguise "amateur reviews". Although consumers are increasingly vigilant, highly disguised false propaganda still leads to a large number of consumers being deceived. Traditional governance solutions have three main limitations: First, the government-led "one-size-fits-all" regulatory model is likely to cause businesses to adopt more covert evasion methods [1, 2], and may suppress the willingness of real users to review due to excessive review [3, 4]. Second, the consumer screening mechanism faces the problem of information asymmetry, and the information transmission solution proposed by Zhang Yang has feasibility obstacles in actual operation [5]. Third, the dynamically evolving false propaganda methods make it difficult for static regulatory measures to be sustained and effective [6].

Existing research has obvious deficiencies in governance thinking. On the one hand, the government's unilateral regulatory model fails to effectively integrate the forces of multiple parties [7]. On the other hand, consumer education programs find it difficult to cope with rapidly changing fraud methods. More importantly, these programs often put the government, consumers, and businesses in opposition and lack the design of a win-win mechanism [8, 9].

Based on signal game theory and AI technology development, this article proposes an innovative four-stage governance plan: In the first stage, a "real evaluation network" is established to solve the problem of "lazy reviews" through a virtual currency incentive mechanism. When the traffic revenue and reputation advantages brought by virtual currency exceed the cost, real evaluation can be effectively incentivized (see Chapter 4 for model derivation). In the second stage, AI dynamic detection is implemented, and technologies such as duplicate image recognition are used to increase the cost of brushing orders, making real evaluation the dominant strategy for merchants. In the third stage, an "authenticity verification button" is set up to open the evaluation function only to actual buyers to improve the credibility of the data. In the fourth stage, a tripartite coordination mechanism

is constructed to form a government-platform-consumer check and balance system through algorithm optimization.

2. Analysis of False Positive Reviews and Description of the Problem of Combating False Positive Reviews

2.1. Analysis of the Causes of False Positive Reviews from the Merchant's Perspective

The essence of merchants choosing fake reviews is the result of a cost-benefit trade-off. The short-term benefits mainly include improving product exposure (platform algorithms tend to recommend highly rated products) and increasing conversion rates (consumers are more inclined to buy well-reviewed products). But they are also triple the cost. First, the cost of returns has increased, especially for non-standard products such as clothing. Second, the crisis of trust was caused by increased consumer vigilance. Third, the risk of impairment of long-term goodwill. In the current market environment, the core reasons for the prevalence of false positive reviews are: first, the platform algorithm relies too much on the number and quality of reviews as the recommendation standard, forming a "bad money drives out good money" effect; second, the supply of real reviews is insufficient, and satisfied users' willingness to review is significantly lower than that of dissatisfied users (the rate of bad reviews is usually 3-5 times that of good reviews); finally, the existing cost of brushing orders (about 5-10 yuan/order) is far lower than the potential benefits.

2.2. Analysis of the Interests of All Parties Involved in Fake Positive Reviews from the Perspective of Other Entities

In the e-commerce evaluation system, the real evaluation mechanism brings significant benefits to multiple parties. Through real reviews, consumers can effectively identify various marketing scams, including commercial promotions disguised as real experiences and misleading behaviors in comment sections manipulated by humans, thereby avoiding purchasing products that do not match the promotion and reducing after-sales disputes. With the help of a real evaluation system, government departments can optimize their regulatory methods, adopt precise subsidies and penalties to replace traditional extensive management, and significantly reduce administrative costs. However, the promotion of real reviews faces resistance from the fake order industry chain and some merchants, who obtain illegal profits through false reviews. Market data shows that consumer complaints caused by false advertising account for 35%, while the implementation costs of traditional regulatory models account for about 15% of the industry size. This conflict of interest has caused the e-commerce evaluation environment to fall into a vicious circle.

2.3. Proposal

To sum up, this plan focuses on expanding the benefits of all parties involved in honest evaluation and reducing the benefits of all parties involved in false evaluation based on the interests of all parties. In the first stage, for merchants, the benefits of honest evaluation are expanded, and the benefits of false evaluation are reduced. This solves the problem of 3.1 in the benefits of telling lies in the merchant's dilemma. However, the problem of lazy evaluation is difficult to solve. The second stage: For consumers, expand the benefits of honest evaluation and ensure their positive initiative and the authenticity of evaluation. At the same time, solve the "lazy evaluation" problem left over from the first stage. The third stage: The government's participation has a positive impact on the overall game situation. Through the three-party game, the three parties cooperate and do not shirk each other. The fourth stage: Use a small amount of effective government means to solve the remaining problems. The government can crack down on the gray industrial chain and reduce the gray income of merchants with false evaluations and consumers who rely on fake orders to make profits.

3. Materials and Method

3.1. Four-stage Specific Solution

Phase 1: Governing issues from the merchant's perspective

To achieve market-oriented governance of real reviews, it is necessary to build a mechanism system that emphasizes both incentives and constraints. In terms of incentives, the compliance costs of honest merchants should be reduced, and their operating income and brand value should be increased through policies such as testing fee subsidies and platform traffic tilting. At the same time, a test report disclosure system should be established to provide consumers with a reliable reference basis, and a spot check mechanism should be used to ensure authenticity. In terms of constraints, the illegal costs of false reviews need to be increased, including credit punishment measures such as mandatory labeling of fake order records and limiting the exposure of problematic products. The seven-day no-reason return system should be improved to maintain consumers' supervision rights. Through differentiated reward and punishment mechanisms, fake order behaviors should be subject to traffic demotion and other measures, forming a market environment where "honesty benefits and dishonesty is restricted", prompting merchants to actively choose the path of real reviews. This precise governance model is more sustainable than traditional supervision [10].

Phase 2: Consumers: Expand the benefits of honest evaluation and increase the benefits of sending real evaluation. At the same time, solve the "lazy evaluation" problem left over from Phase 1.

First, a real evaluation network needs to be built. First, use AI to detect the repetitiveness of evaluation content and pictures. If the same content has the same evaluation in merchant A and merchant B, mark the later good reviews with the word "duplicate" and provide a complaint channel to reduce the risk of misjudgment and injustice; the government has the power to impose fines appropriately. Second, set up an incentive mechanism for consumers to evaluate them truly. It would be a better choice to establish a virtual currency incentive mechanism. The more real evaluations customers generate during consumption, the more virtual currency will be returned. There will be additional virtual currency rewards for more detailed evaluations and those that are recognized for their authenticity. When virtual currency reaches a certain threshold, it can be exchanged for physical money at a 1:1 ratio and punched into the wallet. The specific rules are as follows: Virtual currency rules: 1:1 conversion of virtual currency and physical currency. 1 yuan will be returned for one evaluation.

Phase 3: Use subsequent game models to verify the positive impact of government participation on the overall game situation, and use three-party games to enable the three parties to cooperate with each other and not shirk responsibility.

Phase 4: Using a small amount of government measures to solve the remaining problems: Raising the cost for consumers to send false information

Regarding the problem of false sales in reason 3.1, the government can conduct supervision and prohibit platforms from falsifying sales. Regarding reason 2 in the consumer dilemma, the government can prohibit businesses from hiding negative reviews. When setting the amount of fines, the government can implement phased rewards and punishments, distinguishing fines for small and micro businesses from those for large businesses. It can introduce regulations to crack down on the gray industrial chain and reduce the gray income of businesses that make false reviews and consumers who rely on fake orders to make profits.

3.2. Signaling Game Model

This study constructs a signal game model to analyze the governance mechanism of the e-commerce evaluation system. The model sets market participants as two rational entities, merchants and consumers, and both parties take maximizing benefits as the decision-making goal. Merchants, as signal senders, have private information advantages and can be divided into honest (H) and fraudulent (F) types. Consumers, as signal receivers, cannot directly observe the merchant type and

need to make purchase decisions based on evaluation signals. The government, as a rule maker, provides technical support and institutional guarantees.

In terms of signal cost setting: the signal (SH) of an honest merchant includes real reviews verified by AI and third-party quality inspection reports, with an initial cost of $CH=50$ yuan. The government subsidizes 30 yuan for quality inspection fees and 20 yuan for platform traffic, and the actual cost is reduced to 0 yuan. The signal (SF) of a fraudulent merchant includes forged quality inspection reports and purchase of fake order services, with a total cost of $CF=250$ yuan (forgery cost 150 yuan + evasion detection cost 100 yuan), and an additional 150 yuan fine risk.

The condition for achieving separation equilibrium is: $CF > CH + P$ (i.e. $250 > 0 + 150$). This condition ensures that: fraudulent merchants cannot profit by imitating honest signals; honest merchants gain competitive advantages by relying on low-cost signals; consumers can effectively distinguish between merchant types, forming a market separation equilibrium. The following will verify this equilibrium condition through the game process and the profit function.

3.3. Signal Game Process and Profit Function

Phase 1: Merchant signal sending.

The net profit of honest merchants is $\Pi_H = RH - CH + S = 200 - 50 + 100 = 250$, and honest merchants dominate the market. $RH=200$: sales revenue. $S=100$: platform traffic tilt or government subsidy.) The net profit of fraudulent merchants is $\Pi_F = RF - CF - P = 200 - 250 - 150 = -200$, and they exit the market due to losses. $RF=200$: sales revenue. $P=150$: government fines or platform demotion costs. If you try to imitate the SH signal, the cost is 250 yuan, but the profit is 0 after being detected by AI. Therefore, it can be concluded that honest merchants dominate the market and fraudulent merchants exit at a loss pointed out that signal games can distinguish merchant types under information asymmetry [6]. Verifying that AI technology can quantify signal or cost differences also supports the parameter setting of this game model. The game process is as follows.

Phase 2: Consumer evaluation behavior. Setting up real evaluation incentives (virtual currency mechanism)

First, rule design. For each real review, the buyer will be rewarded 1 yuan of virtual currency. If the consumer review is marked as real, an additional 0.5 yuan will be rewarded. Virtual currency of 50 yuan or more can be exchanged for cash (100% ratio), 30 yuan can be exchanged for 70%, and 10 yuan is guaranteed to be 1 yuan. Unredeemed virtual currency will be cleared every month to stimulate timely evaluation.

Second, behavioral impact. Long-term benefits for consumers (cash + credit) > short-term benefits for fake reviews (such as cashback for fake orders). For example, publishing 50 real reviews will earn 50 yuan in cash and platform membership benefits.

Third, suppression of false reviews (technical support). Use AI detection to mark duplicate content and similar images. Implement device/IP restrictions: The same device can be rated up to 10 times a day to prevent batch fraud. Establish a credit blacklist, and include false reviewers in personal credit files to limit future shopping rights.

3.4. Overall Equilibrium Analysis of Signaling Game

First, the merchant equilibrium result is that the honest merchant (H) dominates the market (profit 250 yuan), and the fraudulent merchant (F) exits due to losses (-200 yuan).

Secondly, in terms of consumer behavior balance, the proportion of real reviews increased: virtual currency incentives + credit rewards. The proportion of false reviews decreased: AI detection + credit penalties.

Perform equilibrium stability analysis. Under the parameterized model, the system equilibrium point is $(x, y) = (1, 0)$, that is, merchants completely choose honest business strategies, and the platform strictly fulfills its detection responsibilities. The basis for achieving this equilibrium is that, first, merchants are given honest incentives. By increasing government subsidy income R_s and fines F_p , the cost of fraudulent behavior is significantly increased ($R_s + C_f + F_p > C_h$), forcing merchants

to switch to honest strategies. Secondly, the platform imposes regulatory constraints. Limiting the additional income R_l of the platform's tacit approval of fake orders can avoid a continuous decline in detection willingness ($C_d + R_l > 0$).

The new solution can bring benefits to three parties. For consumers, by building a "real evaluation network" and a virtual currency incentive mechanism, information asymmetry is reduced, assisting scientific shopping judgment. At the same time, the risk of returns and exchanges is reduced, and the consumer experience is improved. For the government, limited policy intervention (such as subsidies and fines) replaces the traditional "one-size-fits-all" supervision, so that the solution mainly relies on the market, reduces costs, and more efficiently manages false positive reviews; it is conducive to stimulating consumption and responding to the call of national economic policies. For merchants, it is conducive to alleviating the crisis of consumer trust, accumulating business reputation, and increasing the flow of merchants; in order to prevent virtual currency from being cleared, customers will accumulate virtual currency in time, which can promote their consumption in the honest merchant; it can reduce the loss of return and exchange costs.

In summary, by constructing a signal game model, this study proposed an anti-false positive review governance framework based on market-oriented means, aiming to achieve a win-win situation for consumers, the government and businesses.

4. Conclusion

This study found that it is significantly feasible to balance the interests of merchants and platforms in combating false positive reviews. This shows that using market-based means to govern chaos is indeed feasible and more effective. This finding challenges the traditional perception that merchants, platform profits and government market regulation are often in conflict. This also provides us with inspiration for solving similar problems. This can be possibly explained by the following facts. First of all, as consumers' anti-fraud awareness increases, once merchants are found to have false positive reviews, a crisis of trust will arise, leading to customer loss. In addition, under the seven-day no-reason return policy, many consumers will choose to return products if they receive them that do not match the promotion. However, there is a significant secondary sales loss rate for returned goods. As well as increasing packaging and shipping costs, which in turn increases costs. Therefore, choosing to give fake positive reviews is not the Pareto optimal option for merchants. This creates conditions for building a mechanism for synergizing interests. For the platform, building platform credibility is also conducive to creating unique comparative advantages.

The main contribution of this paper is to provide the feasibility of merchants and platforms participating in active governance, which is conducive to better governance. This study first carefully analyzed the traditional solutions, carefully evaluated their advantages and disadvantages, and summarized the feasible and efficient views on this basis, and made appropriate improvements; then, it found two other forces that can be mobilized to solve this problem, so that governance can be transformed from a one-size-fits-all ban to an endogenous driving force for improving market quality. Finally, through the signal game to explore its effectiveness, a more rigorous and feasible and effective measure is provided.

References

- [1] Wu Xiaoli. Why consumers are worried when they can't see negative reviews. China Consumer News, 2024-11-07 (003).
- [2] Zhang Fang, Xu Jingwen. Research on the evolutionary game of e-commerce pre-sale supervision considering false reviews. New Media Research, 2023, 9 (10): 30-36.
- [3] Zhang Yang. Exploring the phenomenon of counterfeiting and shoddy products on e-commerce platforms based on game theory principles. Information Systems Engineering, 2020, (01): 148-149.

- [4] Vincenzo De Angelis, Francesco Buccafurri. A game theory-based approach to discourage fake reviews. *Procedia Computer Science*, 2021, 192: 427-436.
- [5] Prathmesh Shikchi, Shital Jagtap, Jyoti Kundale. Fake review detection using neural network. *ESCI*, 2024, 6: 5-7.
- [6] Wei Dezhi, Chen Fuji, Zheng Xiaoxue. Simulation study on the spread of false information on WeChat based on game theory. *Information Science*, 2016, 34 (11): 146-149+166.
- [7] Edelman Benjamin, Luca Michael. Digital discrimination: The case of Airbnb.com. *Harvard Business School Working Paper*, 2014, 14-054.
- [8] Li Xinxin, Hitt Lorin M. Self-selection and information role of online product reviews. *Information Systems Research*, 2008, 19 (4): 456–474.
- [9] Cabral Luís, Hortacsu Ali. The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics*, 2010, 58 (1): 54–78.
- [10] Tadelis Steven. Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 2016, 8: 321–340.