Adaptive Merchant-Centric Risk Control via Unbiased Decision-Making and Dynamic Optimization in E-Commerce

Xu Liu[1,2*], Yiqiang Lu[2*], Jian Liu[2], Tianyi Zhang[2], Weiqiang Wang[2], Qian Liu[2], Shuai Li[1]

[1] Shanghai Jiao Tong Univerisity, Shanghai, China
[2] Ant Group, Shanghai, China

liu_skywalker@sjtu.edu.cn, yl4025@columbia.edu, rex.lj@antgroup.com, zhangtianyi718@gmail.com, wang.weiqiang@gmail.com, q.liu@wustl.edu, shuaili8@sjtu.edu.cn

Abstract

In the domain of merchant-oriented risk control decisions within e-commerce, balancing the effectiveness of risk management with merchant satisfaction remains a critical challenge. Strict risk control strategies, while effectively mitigating risks, often lead to increased merchant dissatisfaction. Conversely, loose policies could enhance the merchant experience but raise the likelihood of incidents, potentially incurring substantial financial losses. Additionally, determining personalized risk control strategies for different merchants to achieve optimal overall risk management effectiveness is crucial. Given the high uncertainty in the outcomes of different risk control decisions, manual strategy allocation and real-time adjustments are commonly implemented in practice, leading to significant human and resource costs. In this work, we present a novel automated risk control decision framework that utilizes unbiased data-driven decision-making and dynamic optimization to automate the allocation and adjustment of risk control strategies. Our proposed solution adapts to various online business requirements, demonstrating exceptional risk management performance and significantly reducing overall costs. This approach has been extensively deployed and validated in Ant Group's risk control operations, achieving large-scale automated risk control decisions.

Introduction

As digital commerce continues to grow exponentially, the complexity and scale of associated risks escalate, necessitating more sophisticated approaches to risk management (Kim, Ferrin, and Rao 2008). Traditional risk control mechanisms, typically static and rule-based, lack the necessary flexibility and adaptability to effectively manage the dynamic and diverse landscape of merchant activities and external threats. This inadequacy is particularly evident in the fast-paced environment of e-commerce, where decision latency and inflexibility can result in significant financial losses and diminished customer trust (Al-Adwan, Al-Debei, and Dwivedi 2022).

To address these challenges, we introduce an advanced intelligent decision-making framework tailored for the e-commerce sector. This framework leverages the latest advancements in artificial intelligence and machine learning to forge a new path in risk management – one that is dynamic, data-driven, and highly adaptable. Central to our approach is the integration of the contextual bandit algorithm, specifically linUCB, which facilitates real-time, personalized risk management strategies that are both proactive and responsive to the evolving marketplace.

Our framework introduces a robust methodology for generating unbiased estimations of key risk indicators (KRIs) such as the Probability of Risk Occurrence (PRO) and the Probability of Complaint Occurrence (PCO) and learns to optimize from these metrics. Dynamic optimization is another cornerstone of our framework's operational strategy. We designed the system to adjust its risk control measures in real-time, responding adaptively to the fluctuating dynamics of e-commerce activities and merchant profiles. By continuously refining its strategies based on up-to-date data and changing business priorities, our framework adeptly balances risk mitigation with merchant satisfaction, thereby supporting sustainable and safe business operations. The deployment of our framework in real-world settings has yielded compelling results, demonstrating significant improvements over traditional risk management approaches. Moreover, the automation of numerous risk management processes has led to substantial reductions in operational costs and dependencies on manual interventions, enhancing the speed and agility of business decision-making.

The main contributions of this work are as follows:

- We introduce an intelligent decision-making framework for merchant-oriented risk control in e-commerce that is adaptive to diverse business requirements with less cost.
- We integrate data-driven unbiased estimation and math-driven dynamic optimization into the framework that achieves significant performance in application.
- The effectiveness of the framework is widely validated with comparison with baseline methods and online deployment, highlighting its performance and adaptivity in e-commerce platform applications.

Related Work

The research and application of risk control strategies within the realm of e-commerce have witnessed substantial advancements through the integration of machine learning

*These authors contributed equally.

technologies. Historically, many approaches have concentrated on risk identification and analysis through diverse merchant information. For instance, some studies have utilized machine learning techniques to automate the verification of merchant-provided credentials, such as identity documents and business licenses (Laurens and Zou 2016). This automation aids in delineating the risk profiles of different merchants by ensuring the authenticity of submitted information. Additionally, other research has explored the use of temporal features to detect anomalies in merchant transaction data, thereby further characterizing potential risks (Wang and Zhu 2020).

In terms of risk decision-making, several works have aimed at designing and optimizing risk control decisions. This includes the development of expert systems defined by comprehensive risk control workflows (Xia and Chen 2011; Lin et al. 2021), as well as approaches that consider optimizing decision-making from the perspective of uncertainty (Riabacke 2006). These methodologies often involve designing algorithms that can effectively handle uncertain outcomes and multiple objectives within the risk management process (Settembre-Blundo et al. 2021).

Despite these advances, practical applications frequently demand a careful balance between various business metrics according to specific operational needs. This necessitates a system capable of robust risk control or maintaining optimal merchant experience under varying conditions, involving online prediction of multiple metrics and multi-objective optimization (Guo and Zhang 2022; Srinivasan and Kamalakannan 2018). The complexity is further compounded by the massive scale of online data, the delayed nature of feedback, and the intricate relationships among multiple objectives. Moreover, the importance of designing data flows that can efficiently handle and process large-scale data while accounting for delayed feedback is increasingly recognized (Lurie and Swaminathan 2009). Optimizing these data flows ensures that the predictive models and optimization algorithms can operate effectively under the constraints and complexities typical of large e-commerce platforms.

Methodology

In this section, we elaborate on our intelligent decision framework proposed for merchant risk control decisions. We begin by exploring the unique characteristics of problems within the domain of merchant risk control decisions. This includes a detailed description of the data attributes relevant in this scenario and how these characteristics inform the construction of user segmentation and data collection logic, thereby laying the groundwork for our intelligent decision framework. Subsequently, we discuss two key risk indicators (KRIs) of this framework: how to achieve unbiased estimations of the multifaceted impacts of different risk control strategies – including the probability of risk occurrence (PRO) and the probability of complaint occurrence (PCO) – and how to utilize these unbiased estimations for dynamic strategy optimization.

To the best of our knowledge, this is the first intelligent framework in the domain of merchant risk control that leverages data-driven unbiased estimation and mathematically driven dynamic optimization to intelligently allocate and adjust risk control strategies for different merchants to meet diverse business metrics. Our approach has significantly enhanced risk control effectiveness while markedly reducing labor costs and has been successfully applied on a large-scale merchant risk control platform.

Challenges in Risk Control Decision-Making

In this subsection, we delineate the distinct characteristics inherent to the domain of merchant intelligent decision-making within risk control, which sets the stage for our novel framework. The merchant base in contemporary e-commerce platforms is usually vast, often scaling to tens of millions of active users monthly (Erisman 2017). Extensive business insights have achieved the extraction of a series of merchant-specific features such as merchant quantification, transaction records, risk attributes, legal case involvements, control history, and other domain-relevant features. However, the integration of these features for personalized risk control strategy allocation traditionally requires significant manual intervention, such as human-decided rule-based approaches. The choice of risk control strategies brings about varying outcomes; stricter strategies, while preventing potential illicit activities more effectively, tend to deteriorate the merchant experience as evidenced by higher complaint rates. Conversely, more lenient strategies improve merchant satisfaction but increase the likelihood of risk exposure. Manually tailoring these strategies based on experience is not only inefficient but also costly in practice.

Furthermore, in the design of an intelligent system for risk control strategy allocation, it is imperative to consider the unique data characteristics associated with this task. Feedback under risk control strategies is typically delayed (shown in Figure 2). Common types of feedback include risk exposure – indicating the current strategy's failure to prevent risk – and low merchant satisfaction, often resulting from restrictive policies that impair the merchant experience. Typically, there is a time lag, ranging from several days to a month, between the implementation of a risk control strategy and the reception of feedback, necessitating specialized adjustments in data collection and utilization (Jamal et al. 2020; Filippi, Guastaroba, and Speranza 2020).

Evaluating the effectiveness of different risk control strategies also presents a significant challenge. Unlike recommendation systems that strive to maximize user engagement metrics such as click-through rates by recommending products, risk control strategy allocation aims to minimize metrics like the probability of risk occurrence (PRO) and the probability of complaint occurrence (PCO). While recommendation systems employ exploration and exploitation strategies to dynamically learn user preferences through iterative recommendations (Afsar, Crump, and Far 2022), risk control cannot afford similar exploratory learning due to the high costs and potential dissatisfaction from testing different strategies on individual merchants. This limitation inhibits our ability to make unbiased estimations about the effectiveness of unchosen strategies (also known as exposure bias (Chen et al. 2023)), presenting a considerable challenge in the design and implementation of an intelligent system

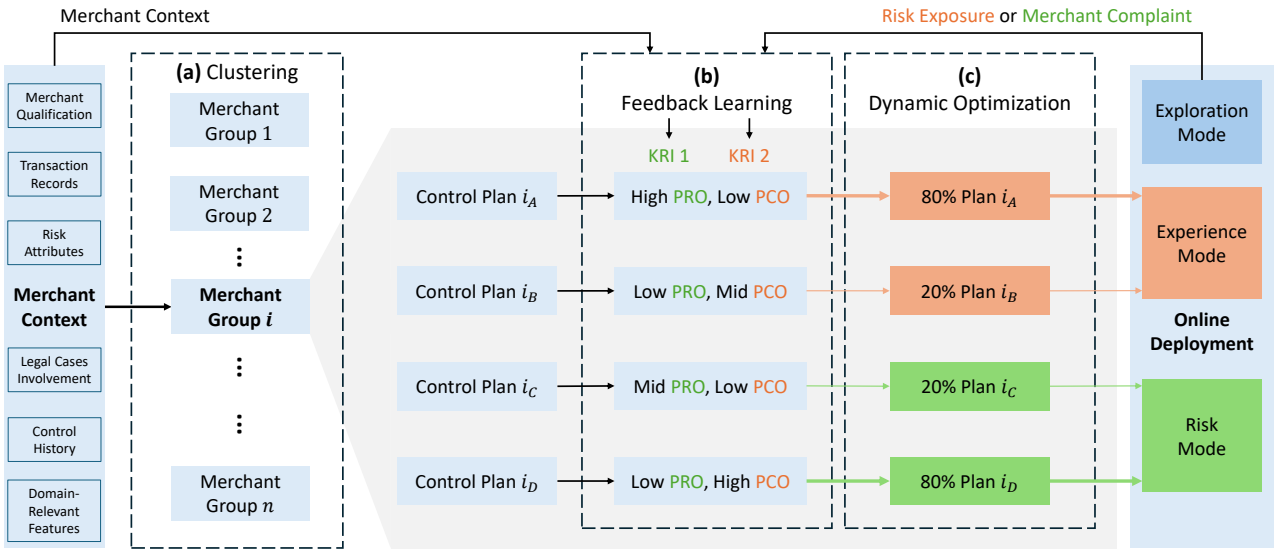

Figure 1: The overview for our merchant-oriented risk control framework. **(a)** Clustering the merchants into different groups with given merchant contexts. **(b)** Feedback learning from delayed signal for predicting two KRIs: PRO and PCO. **(c)** Conduct dynamic optimization for different modes with given target and restrictions to generate final risk management strategy.

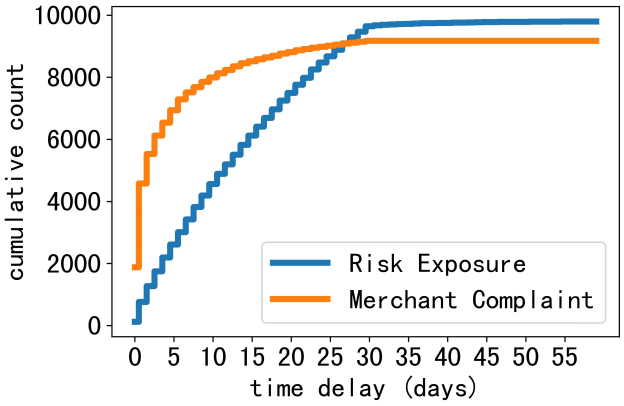

Figure 2: Displaying the accumulation of risk exposures and merchant complaints as functions of time delay, measured in days. The curves highlight the differing rates at which feedback becomes observable, essential for modeling and adjusting risk control strategies based on delayed responses.

capable of unbiasedly assessing various risk control strategies. This challenge underscores the innovative aspect of our framework, which leverages data-driven approaches and dynamic optimization to intelligently allocate and adjust strategies, significantly enhancing the efficiency and effectiveness of merchant risk control. In practical applications, the model account for unbiased estimations across multiple objectives and achieving optimization aligned with business goals, posing challenges in complex large-scale deployment scenarios.

Unbiased Estimation of PRO and PCO

In this subsection, we discuss our approach to achieving unbiased estimations of PRO and PCO, which represent the probability of a merchant experiencing risk exposure or filing a complaint under a specific control action, within our intelligent risk control framework. Central to this approach is the strategic use of merchant smart segmentation and dynamic feedback mechanisms, supported by robust mathematical modeling.

The initial phase involves clustering of merchants. This clustering process leverages merchant-specific features, such as qualifications, transaction volumes, and control history, identical to those previously used in manual risk control strategy decisions. By clustering merchants based on these attributes, our framework facilitates the tailored design and implementation of control measures. The clustering not only mirrors past manual practices but enhances them by allowing dynamic realignment of strategies as merchant risk profiles evolve. For instance, a merchant's change in risk characteristics over time may necessitate their re-allocating into a different group, ensuring that the control measures remain optimally aligned with their current risk profile. In our business practice, there exists a well-established strategy for segmenting merchants, which we denote as $\mathcal{G} = \{G_1, \ldots, G_m\}$, where m represents the total number of merchant groups. For each group G_i, a specific set of candidate risk control actions, denoted by $A_i = \{a_1, \ldots, a_{n_i}\}$, has been designated, with n_i indicating the number of risk control actions available for the i_{th} group. These action sets vary across different groups; for instance, groups categorized under a stricter risk profile tend to have actions more aligned

Group	Available Risk Control Actions
1	Pass Restriction A for 1D Restriction B for 7D Restriction C for 1D Restriction C for 3D
2	Pass Restriction Category A for 1D Restriction Category A for 7D Restriction Category B for 7D Restriction Category B for 14D Restriction Category C for 3D
3	Pass Restriction Category A for 7D Restriction Category A for 14D Restriction Category A for 21D Restriction Category B for 7D Restriction Category B for 14D
4	Restriction Category A for 7D Restriction Category A for 14D Restriction Category A for 21D Restriction Category B for 14D
5	Restriction Category A for 14D Restriction Category A for 21D Restriction Category A for 30D

Table 1: An example for the sets of risk control actions designated for various merchant groups, demonstrating the tailored strategies implemented based on group-specific risk profiles. Actions range from "Pass" (no restriction) to various degrees of payment restrictions. Among these, Category A restrictions are the most stringent, followed by Category B and C. The notation "number + D" in the table indicates that the control action lasts for the specified number of days.

with effective risk management, whereas groups with a more lenient classification lean towards actions that enhance the merchant experience (an example of merchant groups and corresponding risk control action set is shown in Table 1).

In predicting PRO or PCO based on feedback signals, our approach entails estimating the value of the risk control actions corresponding to the current feedback while also considering the characteristics of the merchants on whom these actions are applied. This reflects a critical aspect where the choice of risk control actions impacts PRO or PCO outcomes but does not directly alter the merchant's actual status (e.g., whether they are engaging in illicit activities). Therefore, to evaluate the impact of different decision actions from the context of the merchant's state, we formalize the prediction of PRO and PCO as a contextual bandit problem within our framework, employing a structure similar to the linUCB algorithm for uncertainty estimation (Zhang et al. 2020; Chu et al. 2011).

For instance, consider the prediction of PRO: for a given merchant categorized within group G_i, with an associated set of decision actions $A_i = \{a_1, \ldots, a_{n_i}\}$, we model the reward signal as

$$r = \begin{cases} 1 & \text{if risk occurred} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The predicted reward for a risk control action thus represents the estimated PRO under that action. Due to the constraints on exploration in the risk control context, estimates of PRO for different actions inevitably carry some bias. To better predict uncertainty, we utilize the linUCB algorithm to update the parameters of our prediction function, considering the impact of delayed feedback on model predictions (Chu et al. 2011). Specifically, after obtaining a merchant's feature embedding $\mathbf{x}$, we adjust the learning rate based on the delay time of the feedback received:

$$\alpha = \alpha_0 \cdot \exp(-\gamma \cdot d) \tag{2}$$

where α_0 is a predefined base learning rate, γ is a decay weight for delayed feedback, and d is the time elapsed since the action was taken (measured in days). The estimation of PCO follows a similar methodology. It is important to note that while we adopt the linUCB update strategy (shown in Algorithm 1) to better estimate the potential impact of different actions under given merchant contexts, we do not utilize linUCB's exploration actions for data collection. Due to the practical constraints of online operations that limit extensive proactive exploration, we perform random action selections on a subset of merchants during data collection. We then use the feedback observed under these actions to update the model parameters, thereby enhancing the thorough exploration and accuracy of different action value estimations.

This data-driven approach is underpinned by rigorous safety and production engineering practices to ensure the integrity and reliability of our risk control operations. By integrating advanced mathematical modeling with practical risk control applications, we significantly elevate the decision-making efficacy and optimize the estimation performance for critical metrics like PRO and PCO. This methodology not only refines our understanding of risk dynamics across diverse merchant profiles but also aligns with our overarching goal of deploying an intelligent, adaptable, and efficient risk control system.

Algorithm 1: Estimation Update for PRO or PCO

Require: $\alpha_0, \gamma \in \mathbb{R}_+$, observation for merchant and action context $\mathbf{x} \in \mathbb{R}^k$, reward r and delay d on action a

if a has not been observed before **then**
  $\mathbf{A}_a \leftarrow \mathbf{I}_k$ (k-dimensional identity matrix)
  $\mathbf{b}_a \leftarrow \mathbf{0}_{k\times 1}$ (k-dimensional zero vector)
end if
$\alpha \leftarrow \alpha_0 \cdot \exp(\gamma \cdot d)$ (equation (2))
$\mathbf{A}_a \leftarrow \mathbf{A}_a + \mathbf{x}\mathbf{x}^\top$
$\mathbf{b}_a \leftarrow \mathbf{b}_a + r_t\mathbf{x}$
Value estimation $\leftarrow \left(\mathbf{A}_a^{-1}\mathbf{b}_a\right)^\top \mathbf{x}$
Optimistic $\leftarrow \left(\mathbf{A}_a^{-1}\mathbf{b}_a\right)^\top \mathbf{x} + \alpha_0 \cdot \sqrt{\mathbf{x}^\mathbf{T}\mathbf{A}_a^{-1}\mathbf{x}}$

Adaptive Optimization of Risk Control Strategies

In this subsection, we detail the methods for dynamically optimizing risk control strategies upon obtaining reliable estimations for PRO and PCO. In practical business scenarios, it is essential to strike a balance between controlling risk incidents and maintaining merchant satisfaction. Specifically, our strategy entails maximizing core metrics while adhering to overarching constraints, aiming to optimize risk probabilities without compromising user experience, or alternatively, to enhance merchant experience without increasing risk probabilities. This adaptive approach allows us to tailor the intelligent framework's risk control strategies to the diverse business needs of different merchants (Gunantara 2018).

Our dynamic decision system is designed with three operational modes based on business decision preferences (shown in Figure 3). The first mode, as previously described, is the exploration mode used for unbiased data collection, where risk control actions are randomly selected to provide unbiased training data essential for optimizing estimation algorithms. The second mode focuses on risk management, aiming to optimize PRO while ensuring that the rate of merchant complaints does not exceed that observed under manual risk control strategies. The third mode, centered on enhancing the user experience, seeks to improve PCO under the condition that the probability of merchant risk occurrence remains at or below the levels managed by manual strategies. The proportions of these three modes can be dynamically adjusted according to the overall control style required by the business, and constraints can be modified as needed to specify more flexible control plans based on evolving business demands.

For the risk and experience modes, we employ mathematically driven operations research to further optimize the allocation of risk control strategies. Taking the risk management mode as an example, we can formulate the problem mathematically as an optimization problem:

$$\begin{aligned} \min \quad & \left(\sum_{ij} \mathrm{risk}_{ij} \cdot a_{ij} \right) \\ \text{s.t.} \quad & \sum_{j} a_{ij} = 1, \quad a_{ij} \in \{0,1\} \\ & \sum_{ij} \mathrm{exp}_{ij} \cdot a_{ij} \leq c_{\text{initial}} \end{aligned} \tag{3}$$

where each merchant is denoted by i and each risk control strategy (a_{ij}) applicable to the merchant is represented by j. The objective of the optimization is to minimize the overall PRO for all merchants, and the PRO for merchant i under strategy j is denoted as risk_{ij}, as estimated using the methods described in the previous subsection.

This optimization problem incorporates two types of constraints. The first constraint ensures operational feasibility by mandating that exactly one risk control strategy is applied to each merchant, which guarantees that the optimized strategy allocation can be practically implemented. The second constraint pertains to the merchant experience, stipulating that the predicted PCO under the optimized strategy allocation, denoted as exp_{ij}, should not exceed the PCO levels achieved prior to optimization, represented by c_{initial}. Initially, the initial PCO level is calculated based on manually defined risk control allocations from previous work; subsequently, it is based on the outcomes of the last optimization cycle. This iterative adjustment allows for continuous optimization of risk control strategy allocation within our system, ensuring that adjustments not only respond to changing conditions but also progressively refine the balance between managing risk and preserving merchant satisfaction. By integrating the online primal-dual algorithm (Buchbinder and Naor 2009), the Lagrange multiplier λ^* can be represented as

$$\begin{aligned} \lambda^* &= \arg\max_{\lambda \geq 0} \inf_{a \in A} L(a, \lambda) \\ L(x, \lambda) &= \sum_{ij} \mathrm{risk}_{ij}\, a_{ij} + \lambda \Big(\sum_{ij} \mathrm{exp}_{ij}\, a_{ij} - c_{\text{initial}} \Big) \\ A &= \big\{ a_{ij} \big| \sum_{j} a_{ij} = 1, \forall i \big\} \end{aligned} \tag{4}$$

and at inference time, for new-coming merchant i, the risk management strategy can be decided by

$$a_{ij} = \begin{cases} 1, & \text{if } j = \arg\max_j (-\mathrm{risk}_{ij} - \lambda^* \mathrm{exp}_{ij}) \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The formalization of the optimization problem for the experience mode follows a similar methodology to that of the risk mode, with a key alteration in the optimization objective. In this mode, the primary goal is to minimize PCO, reflecting our focus on enhancing merchant satisfaction. While adhering to the constraint that the number of strategies applied must remain consistent, we also ensure that the overall PRO does not exceed the levels observed prior to optimization. This sophisticated approach ensures that our risk control framework remains both dynamic and responsive to the real-time needs of the business, fostering an environment where risk control and merchant satisfaction are harmoniously balanced.

In summary, our approach utilizes three strategic modes – exploration, risk, and experience – to dynamically manage merchant risk. The exploration mode underpins our system, facilitating unbiased data collection through random selection of risk control actions, ensuring data free from selection and exposure biases. The risk mode aims to optimize PRO while maintaining merchant satisfaction within acceptable limits, effectively balancing risk mitigation with positive merchant experiences. The experience mode, conversely, focuses on minimizing PCO to boost merchant satisfaction, ensuring that risk levels do not exceed predefined thresholds. Collectively, these modes allow for a dynamic and adaptive approach to risk control strategy allocation. They provide a methodological framework that supports continuous improvement through iterative optimization cycles, each informed by the latest data and performance metrics. The flexibility to adjust the balance between risk mitigation and merchant satisfaction according to real-time data and evolving business needs is a significant advantage of our system.

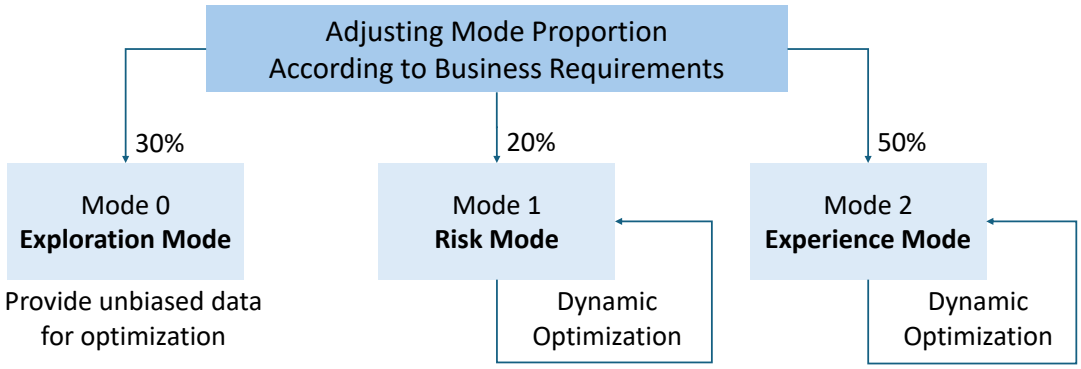

Figure 3: An example of the distribution and dynamic adjustment of three operational modes used in our risk control framework. Each mode is designed to address specific objectives: Exploration Mode (30%) provides unbiased data essential for model optimization, Risk Mode (20%) focuses on dynamic optimization PRO while maintaining merchant satisfaction, and Experience Mode (50%) prioritizes enhancing the merchant experience through dynamic optimization of PCO. The proportions of these modes are adjustable based on evolving business requirements and diverse operational needs.

Experimental Evaluation

Experimental Objectives

The primary goal of our experimental evaluation is to rigorously test the effectiveness and efficiency of our novel intelligent risk control framework. Specifically, we aim to validate the precision of our models in estimating PRO and PCO. By comparing these models against traditional benchmarks in risk assessment, we seek to highlight the advantages of our data-driven, adaptive approach in terms of accuracy and contextual responsiveness. Besides, we explore the efficiency and economic viability of implementing our automated system compared to manual methods and some alternative approaches such as using linUCB's optimistic predictions for optimizing single metrics or employing Pareto front for multi-objective optimization (Giagkiozis and Fleming 2014). The goal from these comparisons is to illustrate that our dynamic optimization approach not only outperforms in terms of performance but also excels in applicability, controllability, and computational efficiency.

Experiment Setup

In the setup phase of our experimental evaluation, we leveraged the extensive experience of our business teams in risk management to categorize merchants requiring risk control into six distinct groups. Each group was assigned a set of possible control actions, ranging from three to ten options per group. This clustering facilitates targeted and efficient management across varied risk profiles. During the model training phase for unbiased estimation of the PRO and PCO, we adopted an online training approach. Specifically, merchant feedback data collected daily through an exploratory mode were utilized to continuously update our models. To ensure the robustness of our results, 30% of the merchants were assigned to this exploratory mode, with the data collection period spanning from January 10, 2024, to May 10, 2024, ensuring consistent data availability across different methodologies tested.

The deployment of our models was strategically aligned with business needs, where approximately 20% of the merchants were managed under the risk mode, about 50% under the experience mode, and the remaining in the exploratory mode. This distribution was designed to optimize the overall performance of the system, adjusting parameters daily based on outcomes. For the unbiased estimations, we set the learning rate parameter α_0 at 1.0 and the decay parameter γ at 0.01. Each merchant group was subjected to two predictive models estimating PRO and PCO for the actions within the group. The accuracy of these predictions was evaluated using the ROC-AUC score (Yang and Ying 2022). Once the predictive models were established, their outputs facilitated optimistic estimations used in calculating risk and experience functions within our dynamic optimization framework. The resulting strategies were then deployed to merchants, with ongoing feedback collected to further refine and update the models.

To assess the efficacy of our framework, particularly its superiority in real-world applications, we conducted a comparative analysis between the automated strategies generated by our framework and traditional manual strategies. Merchants were evenly split into two groups: one managed by our new framework and the other continuing with manual risk control approaches. This setup allowed for a direct comparison over a one-month period, during which feedback was collected to evaluate the performance improvements brought by our automated system. Moreover, to gauge the effectiveness and efficiency of our dynamic optimization approach, we also compared it against other baseline meth-

Approach	Average PRO	Relative PRO Reduction	Average PCO	Relative PCO Reduction
Manual Strategy	0.44	0%	0.20	0%
Single (PRO Minimization)	0.11	-75%	0.72	+260%
Single (PCO Minimization)	0.58	+32%	0.02	-90%
Pareto Front	0.39	-11%	0.26	+30%
Our Approach (Risk Mode)	0.34	-23%	0.20	0%
Our Approach (Experience Mode)	0.44	0%	0.09	-55%

Table 2: Comparative analysis of risk control strategies illustrating average PRO and PCO values alongside relative reductions achieved by our approach and the baseline approaches. The risk mode and Experience mode in our approach can effectively reduce the PRO or PCO without sacrificing the other measure.

ods. These included strategies derived solely from linUCB's optimistic estimates aimed at minimizing PRO or PCO, as well as strategies calculated from a Pareto front (Hua et al. 2021) to achieve a balanced optimization between the two metrics.

Experimental Results

In our experimental evaluation, the predictive accuracy of our models for PRO and PCO achieved remarkable results, evidenced by AUC-ROC scores of **0.87** and **0.84**, respectively. This achievement was largely due to our data flow design which incorporates random exploration for extensive data collection and optimistic exploration for uncertainty estimation. This robust data collection strategy enabled our models to effectively capture the complex dynamics of merchant behaviors and risk patterns, thereby enhancing the reliability of our predictions.

The efficacy of our dynamic optimization framework is particularly noteworthy. By enabling the selection between risk mode and experience mode, our system adeptly balances the dual objectives of minimizing risk without compromising merchant satisfaction, and vice versa. According to the results summarized in Table 2, when operating under the risk mode, our framework facilitated a 23% reduction in PRO while maintaining merchant satisfaction levels. Conversely, under the experience mode, it achieved a 55% reduction in PCO without increasing risk. This capability to optimize one metric without adversely affecting the other underscores the unique advantage of our method: it adapts flexibly to varying business needs while significantly reducing the need for manual intervention, thus lowering operational costs.

Moreover, our approach was benchmarked against several alternative baselines, including those focused on single-objective optimizations of PRO or PCO. These comparisons revealed that while optimizing one of these conflicting metrics often results in the deterioration of the other, our method excels in balancing both. This contrast is stark, especially when compared with strategies that use a Pareto front for multi-objective optimization. While the intuition behind using a Pareto front is to find a compromise between conflicting objectives, it often fails to achieve a practical balance in real-world applications, typically resulting in increased PCO. In contrast, our framework not only provides more flexible outcomes but also operates with significantly greater efficiency. During deployment and testing, the time required to make a round of risk control decisions in our system was only one-fifth of that needed for calculating strategies based on the Pareto front. This efficiency gain is particularly valuable in real-time decision-making scenarios involving a large number of merchants, underscoring the practical benefits of integrating unbiased estimation with dynamic optimization in our risk control framework.

Application Use and Payoff

In deploying our intelligent risk control framework within Ant Group, we managed approximately 16,000 merchants daily. The framework dynamically allocated risk control strategies to these users and continuously optimized decisions based on real-time feedback. Over time, this system demonstrated a substantial positive impact, achieving a significant reduction in overall risk rates and merchant complaints. Specifically, the risk exposure rate decreased by 33%, and merchant complaints dropped by 69%, marking a notable improvement in both risk mitigation and user satisfaction. Additionally, the model was computationally optimized during deployment at Ant Group, ensuring that the introduced computational demand remained acceptable in high-traffic e-commerce environments.

One key element of this success was the implementation of periodic monitoring and reporting. By analyzing merchant risk exposure rates and complaint rates over 30-day cycles, the system dynamically adjusted the traffic allocation proportions among risk mode, exploration mode, and experience mode. These adjustments ensured that the framework remained agile in addressing evolving business conditions while maintaining operational efficiency. However, early deployment revealed challenges. Initial results showed that the risk mode did not effectively control risk exposure rates, and the model tended to favor lenient strategies. Analysis suggested this was due to a higher proportion of weak control actions in the training samples gathered during the exploration phase, leading to inflated confidence in their effectiveness. To address this, we temporarily increased the proportion of traffic allocated to exploration mode. This change yielded more diverse data, enabling the model to better estimate the value of different risk control strategies.

The adaptive and automated nature of the framework further integrated seamlessly with real-time monitoring and alert systems. These systems tracked decision outputs, merchant segmentation, and operational anomalies, ensuring the stability and security of large-scale e-commerce operations.

This continuous oversight, coupled with the ability to adjust traffic proportions dynamically, allowed the system to adapt effectively to market demands, ensuring consistent alignment with business objectives.

Lessons Learned

The deployment of our proposed intelligent risk control framework in Ant Group's large-scale e-commerce environment has highlighted the following key lessons:

1. One of the most significant deployment challenges was ensuring the diversity and representativeness of training data, particularly during the model's exploratory phase. The merchant grouping based on past business experience provides valuable support but may also introduce potential biases. Early observations revealed that the model's decisions were overly biased toward lenient control strategies, leading to suboptimal risk mitigation. This issue was traced back to the disproportionate prevalence of weak control actions in the initial exploration samples, which inflated the model's confidence in their effectiveness. To address this, we increased the proportion of traffic allocated to the exploration mode. This adjustment ensured a more comprehensive sampling of merchant behaviors across different control strategies. Over time, the framework's decision-making improved, demonstrating a more balanced allocation of diverse actions.

2. The inherent delay in receiving feedback from merchants, such as risk exposures or complaints, posed a significant challenge to real-time model updates. For example, risk exposure events or complaints might occur days or weeks after a control action is applied, creating a temporal gap that, if not addressed, could lead to misleading estimations and suboptimal model performance. Our solution integrated delayed feedback into the contextual bandit algorithm using time-adjusted learning rates. The model adapted its confidence in action-value predictions based on the delay duration, ensuring that recent but delayed outcomes were still factored into decision-making. This mechanism improved the accuracy of estimations for key risk indicators (PRO and PCO) despite the feedback delays. The experience emphasized that addressing delayed feedback requires not only algorithmic sophistication but also robust engineering workflows that align the data processing pipeline with real-world operational timelines.

3. The ability to dynamically adjust traffic proportions across operational modes proved essential for maintaining system effectiveness in different business scenarios. For instance, during high-risk periods, such as major shopping holidays, the risk mode required a larger traffic share to handle heightened fraud risks. Conversely, in stable periods, a higher allocation to exploration and experience modes allowed the model to refine its long-term learning and improve merchant satisfaction. These adjustments were informed by periodic monitoring reports that analyzed merchant risk exposure rates and complaint rates over 30-day cycles. By leveraging this real-time data, we proactively tailored the framework's priorities to align with business objectives.

Conclusion and Future Work

Conclusion

In this study, we introduced a novel intelligent risk control framework designed to optimize risk management strategies dynamically within an e-commerce context. Our approach successfully integrates unbiased estimation of risk probabilities with dynamic optimization to offer a flexible, robust solution adaptable to various merchant needs. The experimental results substantiated the effectiveness of our models, demonstrating substantial improvements in both PRO and PCO across different operational modes. Specifically, our framework achieved a significant reduction in PRO by 23% and PCO by 55% in respective risk and experience modes compared to traditional manual strategies. These outcomes not only highlight the efficiency of our approach in balancing risk mitigation and merchant satisfaction but also underscore its potential in reducing operational costs by minimizing the need for manual interventions. Furthermore, the real-time monitoring capabilities embedded within our system ensure ongoing oversight and the ability to respond swiftly to anomalies, enhancing both the safety and stability of the decision-making process in real-world applications.

Limitations and Future Work

While our framework advances automated risk control, it has limitations that merit further exploration. A key challenge is its reliance on extensive, diverse, and accurate data, which may restrict its applicability in data-constrained or biased scenarios. Future work could extend the framework to incorporate multi-dimensional risk factors for more comprehensive assessments and decision-making across diverse business contexts. We also aim to enhance decision explainability by developing mechanisms to provide clear, interpretable justifications for risk control decisions, fostering trust and usability. Additionally, we plan to create a more agile risk control system with intuitive tools for real-time strategy adjustment, enabling dynamic responses to evolving business needs and market dynamics in e-commerce.

Acknowledgements

The authors from Ant Group are supported by the Leading Innovative and Entrepreneur Team Introduction Program of Hangzhou (Grant No.TD2022005).

References

Afsar, M. M.; Crump, T.; and Far, B. 2022. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7): 1–38.

Al-Adwan, A. S.; Al-Debei, M. M.; and Dwivedi, Y. K. 2022. E-commerce in high uncertainty avoidance cultures: The driving forces of repurchase and word-of-mouth intentions. *Technology in Society*, 71: 102083.

Buchbinder, N.; and Naor, J. 2009. Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research*, 34(2): 270–286.

Chen, J.; Dong, H.; Wang, X.; Feng, F.; Wang, M.; and He, X. 2023. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, 41(3): 1–39.

Chu, W.; Li, L.; Reyzin, L.; and Schapire, R. 2011. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 208–214. JMLR Workshop and Conference Proceedings.

Erisman, P. 2017. *Six billion shoppers: the companies winning the global e-commerce boom*. St. Martin's Press.

Filippi, C.; Guastaroba, G.; and Speranza, M. G. 2020. Conditional value-at-risk beyond finance: a survey. *International Transactions in Operational Research*, 27(3): 1277–1319.

Giagkiozis, I.; and Fleming, P. J. 2014. Pareto front estimation for decision making. *Evolutionary computation*, 22(4): 651–678.

Gunantara, N. 2018. A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1): 1502242.

Guo, K.; and Zhang, L. 2022. Multi-objective optimization for improved project management: Current status and future directions. *Automation in Construction*, 139: 104256.

Hua, Y.; Liu, Q.; Hao, K.; and Jin, Y. 2021. A survey of evolutionary algorithms for multi-objective optimization problems with irregular Pareto fronts. *IEEE/CAA Journal of Automatica Sinica*, 8(2): 303–318.

Jamal, A.; Tauhidur Rahman, M.; Al-Ahmadi, H. M.; Ullah, I.; and Zahid, M. 2020. Intelligent intersection control for delay optimization: Using meta-heuristic search algorithms. *Sustainability*, 12(5): 1896.

Kim, D. J.; Ferrin, D. L.; and Rao, H. R. 2008. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems*, 44(2): 544–564.

Laurens, R.; and Zou, C. C. 2016. Using credit/debit card dynamic soft descriptor as fraud prevention system for merchant. In *2016 IEEE Global Communications Conference (GLOBECOM)*, 1–7. IEEE.

Lin, S.-S.; Shen, S.-L.; Zhou, A.; and Xu, Y.-S. 2021. Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. *Automation in Construction*, 122: 103490.

Lurie, N. H.; and Swaminathan, J. M. 2009. Is timely information always better? The effect of feedback frequency on decision making. *Organizational Behavior and Human decisión processes*, 108(2): 315–329.

Riabacke, A. 2006. Managerial Decision Making Under Risk and Uncertainty. *IAENG International Journal of Computer Science*, 32(4).

Settembre-Blundo, D.; González-Sánchez, R.; Medina-Salgado, S.; and García-Muiña, F. E. 2021. Flexibility and resilience in corporate decision making: a new sustainability-based risk management system in uncertain times. *Global Journal of Flexible Systems Management*, 22(Suppl 2): 107–132.

Srinivasan, S.; and Kamalakannan, T. 2018. Multi criteria decision making in financial risk management with a multi-objective genetic algorithm. *Computational Economics*, 52(2): 443–457.

Wang, C.; and Zhu, H. 2020. Representing fine-grained co-occurrences for behavior-based fraud detection in online payment services. *IEEE transactions on dependable and secure computing*, 19(1): 301–315.

Xia, D.; and Chen, B. 2011. A comprehensive decision-making model for risk management of supply chain. *Expert Systems with Applications*, 38(5): 4957–4966.

Yang, T.; and Ying, Y. 2022. AUC maximization in the era of big data and AI: A survey. *ACM computing surveys*, 55(8): 1–37.

Zhang, X.; Xie, H.; Li, H.; and CS Lui, J. 2020. Conversational contextual bandit: Algorithm and application. In *Proceedings of the web conference 2020*, 662–672.