# Aggregation Rules and Dynamic Weight Optimization in Multi-Source Evaluation Mechanisms: An Institutional Analysis of Ranking Robustness in Competitive Markets

Shuo Zhang *

School of International Economics, China Foreign Affairs University, Beijing, China

* Corresponding Author Email: 1733091697@qq.com

**Abstract.** This paper addresses the issues of weight structure and aggregation rules in hybrid evaluation mechanisms by constructing an analytical framework that combines backward inference with dynamic optimization. Under conditions of incomplete observational data, the “minimum necessary support” model is introduced. Combined with Monte Carlo simulations to characterize the feasible space of unobservable variables, this approach reveals the structural amplification effect of evaluation source differences under percentage summation rules. Further comparative analysis demonstrates that ranking methods effectively constrain extreme concentration shocks, enhancing result robustness. Building upon this, a multi-objective dynamic weight optimization framework is proposed. It identifies balanced weight intervals via the Pareto frontier and designs a phased dynamic adjustment mechanism to harmonize early-stage responsiveness with late-stage stability. This research integrates reverse inference, distribution simulation, and optimization analysis to provide a universal quantitative basis for designing multi-source information weighting and decision aggregation mechanisms.

**Keywords:** Evaluation Mechanism Design, Preference Concentration Effects, Weight Optimization.

## 1. Introduction

In competitive markets involving multiple participants, a central challenge in resource allocation and outcome formation is the effective integration of information from diverse sources. In practice, evaluation systems often combine expert scoring and public preferences, which differ substantially in information structure, volatility, and concentration. Under simple aggregation rules, these distributional disparities may be amplified, reshaping ranking structures and potentially weakening the stability and incentive effects of final outcomes. Therefore, evaluation rules are not merely technical arrangements but institutional variables influencing behavior and market expectations.

Existing studies mainly focus on optimizing single evaluation criteria or examining sensitivity under fixed weighting schemes, with limited institutional comparisons of how alternative aggregation rules transmit volatility and induce ranking reversals. In weighting design, fixed proportional allocation remains common, overlooking stage-specific dynamics and incentive adjustments. Moreover, unobservable preference variables are often treated with oversimplified assumptions, failing to reflect realistic distribution patterns. These limitations make evaluation mechanisms vulnerable to concentrated preferences or extreme shocks [1].

To address this, this paper develops an analytical framework integrating reverse inference, distribution simulation, and multi-objective optimization, grounded in rule structures and distributional features. It first characterizes the feasible space of unobservable variables under incomplete information, then compares summation-based and ranking-based mechanisms in terms of volatility constraints and risk mitigation. Furthermore, dynamic weighting functions are introduced to quantify the trade-off between fairness and activity, with frontier analysis identifying balanced intervals [2]. The main contribution is transforming rule design into a measurable and optimizable mechanism configuration problem, providing systematic tools for institutional arrangements in multi-source information environments.

# 2. Support Inference Framework under Incomplete Information

## 2.1. Modeling the Aggregated Evaluation Mechanism

In this section, a mathematical framework is established to analyze the scoring and elimination mechanisms, and to quantitatively evaluate the potential impact of fan votes on competition outcomes. Given that specific fan vote data are not publicly available, the core challenge lies in reverse-engineering the distribution characteristics of fan votes and their weighting effects in specific seasons using known judges' scores and elimination results.

Let $n_t$ denote the number of contestants competing in week $t$ of a given season. For contestant $i$ in week $t$: $J_{i,t}$ represents the total score awarded by judges; $V_{i,t}$ represents the audience vote count (unobserved).

Season 17 adopts the Percent Combined method. Under this system, the judges' score percentage and the audience vote percentage are defined as:

$$P_{J,i,t} = \frac{J_{i,t}}{\sum_{k=1}^{n_t} J_{k,t}} \times 100,\ P_{F,i,t} = \frac{V_{i,t}}{\sum_{k=1}^{n_t} V_{k,t}} \times 100 \tag{1}$$

The total score is defined as:

$$S_{i,t} = P_{J,i,t} + P_{F,i,t} \tag{2}$$

The elimination rule states that the contestant with the lowest total score is eliminated. Therefore, for any eliminated contestant $e$ and any surviving contestant $j$, the following inequality must hold $S_{e,t} \le S_{j,t}$. This inequality forms the core constraint for inferring audience voting behavior $P_{F,i,t}$. Since $V_{i,t}$ remains unobserved and the number of unknown variables exceeds the number of equations, the system is underdetermined. Rather than identifying a unique solution, the objective is to characterize the feasible solution space.

To address this issue, a Minimum Required Fan Support (MRFS) model is introduced. This model is designed to calculate the minimum required percentage of audience votes for a low-scoring contestant to avoid elimination under the observed judges' scores and survival threshold.

## 2.2. Simulation-Based Inference of Feasible Vote Distributions

Model validation is conducted using Season 17 as a case study. Judges' scoring data for each week are extracted and combined with the actual elimination list to perform reverse inference of feasible audience vote distributions.

The solution procedure includes the following steps:

1. Data normalization

Weekly judges' scores are converted into percentage form using:

$$P_{J,i,t} = \frac{J_{i,t}}{\sum_{k=1}^{n_t} J_{k,t}} \times 100 \tag{3}$$

This ensures comparability across contestants within each week.

2. Threshold determination

For each elimination round, a survival threshold is established. To survive, a contestant's total score must exceed that of the lowest-scoring surviving contestant. Let $S_{\min,t}$ denote that benchmark.

Then $S_{i,t} > S_{\min,t}$. This condition generates inequality constraints on the unobserved audience vote share.

3. Monte Carlo simulation

Given the underdetermined nature of the equation system, Monte Carlo simulation is employed to generate admissible audience vote distributions [3]. Consistent with empirical patterns observed in entertainment markets, audience votes are assumed to follow a skewed long-tail distribution (e.g., Zipf-type distribution). Random samples of $V_{i,t}$ are generated and retained only if all historical elimination inequalities are satisfied.

4. Feature extraction

Within the feasible solution space, the mean audience vote proportion of contestants characterized by low judges' scores but strong popularity is calculated. This provides an estimate of the minimum vote share required for technically weak contestants to survive.

### 2.3. Results Analysis

In the Fig, 1, the results reveal that audience voting plays a significant role under the percentage-based aggregation system, particularly for contestants positioned in the mid-to-low range of judges' evaluations. Because judges' scores are typically concentrated within a narrow interval (e.g., 7–9 points), the variance of $P_{J,i,t}$ remains limited. In contrast, audience votes may exhibit substantial dispersion due to popularity concentration.

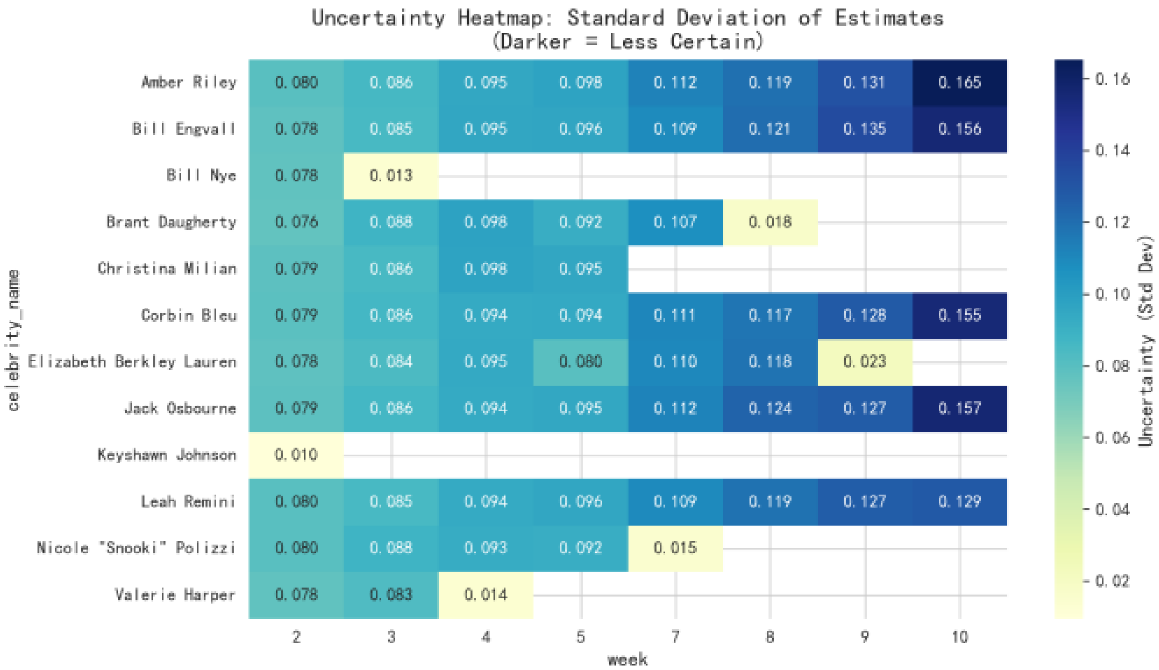


**Figure 1.** Season 17 fan vote model analysis

Under the additive aggregation rule, the relative impact of each component depends on its dispersion. When the variability of $P_{F,i,t}$ exceeds that of $P_{J,i,t}$, audience influence becomes structurally amplified within the scoring mechanism.

This structural amplification implies that concentrated audience support can offset observable differences in expert evaluations. In such a setting, advancement outcomes are jointly determined by professional quality signals and the distributional intensity of audience preferences. Consequently, the percentage-based rule operates as a hybrid allocation mechanism in which both expert assessments and consumer demand contribute to competitive equilibrium.

This institutional characteristic provides the analytical foundation for examining alternative aggregation rules and their potential redistributive effects on competitive outcomes.

## 3. Sensitivity and Institutional Robustness of Aggregation Rules

This section establishes a comparative model to quantify the sensitivity differences between the Percentage Method and the Ranking Method when integrating heterogeneous evaluation sources, namely objective judges' scores and subjective audience votes. The objective is to evaluate how alternative aggregation rules respond to distributional asymmetry and to assess the institutional rationale behind rule adjustments.

### 3.1. Institutional Differences between Percentage and Ranking Rules

Under the percentage system, for a week with $N$ contestants, the total score of contestants $i$ is $S_{pct,i} = P_{J,i} + P_{F,i}$, where $P_{J,i}$ and $P_{F,i}$ denote the percentage scores from judges and audience, respectively.

Empirically, judges' scores typically cluster within a narrow interval (e.g., 7–10 points), implying relatively low variance $\sigma_J^2$. In contrast, audience vote distributions often exhibit heavy-tailed characteristics (e.g., Pareto-type concentration), leading to higher variance $\sigma_F^2$.

Under the additive aggregation structure, total variance satisfies:

$$\mathrm{Var}(S_{pct,i}) = \sigma_J^2 + \sigma_F^2 + 2\mathrm{Cov}(P_{J,i}, P_{F,i}) \tag{4}$$

When $\sigma_F^2 \gg \sigma_J^2$, total score variability is primarily driven by audience votes. According to weighted-sum variance decomposition, the component with greater dispersion exerts disproportionate influence. Consequently, the percentage system structurally amplifies popularity shocks, granting audience votes de facto veto or promotion power.

In the ranking system, the total score is defined as:

$$S_{rank,i} = R_{J,i} + R_{F,i}, \text{where } R_{J,i}, R_{F,i} \in 1, 2, \dots, N \tag{5}$$

Here, cardinal utilities (raw scores) are converted into ordinal utilities (rankings). Because the gap between adjacent ranks is fixed at one unit, variance contributions from judges and audience are mechanically bounded. This transformation reduces sensitivity to extreme vote concentration and equalizes dispersion across evaluation sources.

### 3.2. Model Solution

A counterfactual evaluation is conducted using the Season 27 finals, where Bobby Bones won despite consistently weak judges' assessments. This case provides a clear setting to compare how aggregation rules transmit popularity advantages into final outcomes.

Under the percentage-based rule, judges' scores are normalized within a narrow range. For example, if the finalists' raw judges' scores are approximately $\{80, 75, 78, 55\}$, then Bobby's judges' percentage is:

$$P_{J,B} = \frac{55}{80 + 75 + 78 + 55} \approx 0.19, \tag{6}$$

While the leading contestant achieves $P_{J,M} \approx 0.28$. The technical gap is therefore less than 10 percentage points after normalization. Such compression implies that a moderate advantage in audience voting can overturn expert-based rankings, particularly when votes follow a highly concentrated distribution.

In contrast, under the ranking-based rule, Bobby's lowest judges' score directly translates into last place $R_{J,B} = 4$, while the technical leader obtains $R_{J,M} = 1$.

Even if Bobby ranks first in popularity ( $R_{F,B} = 1$ ) and the opponent remains mid-ranked ( $R_{F,M} = 3$ ), the total scores become $S_{rank,B} = 5$, $S_{rank,M} = 4$.

Thus, ordinal aggregation limits the reversibility of technical deficits, preventing extreme popularity shocks from fully dominating outcomes.

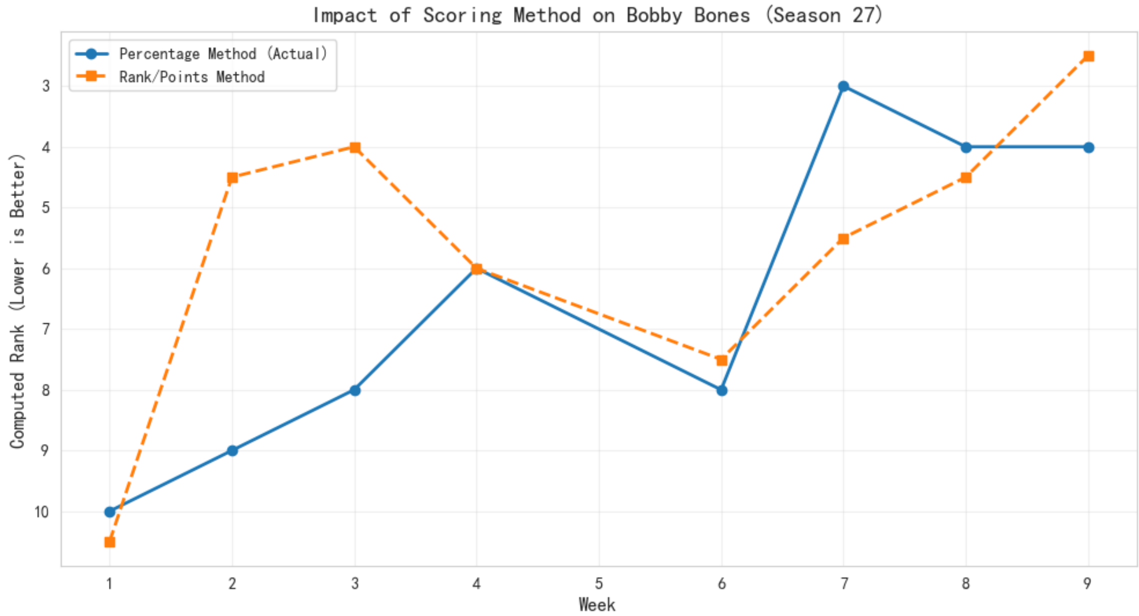


**Figure 2.** Method comparison Bobby Bones

Fig. 2 illustrates that under the percentage-based aggregation rule, Bobby Bones' ranking improves in several weeks despite weaker judges' evaluations, indicating that concentrated audience support can offset technical disadvantages within an additive percentage structure. In contrast, the ranking-based simulation keeps his computed position consistently lower, since ordinal aggregation bounds the marginal influence of popularity and limits reversals of professional ordering. This divergence highlights that aggregation rules fundamentally shape how audience preference shocks are transmitted into advancement outcomes, with the percentage system amplifying demand-side concentration while the ranking rule provides greater institutional robustness.

The comparative analysis demonstrates that the ranking system reduces sensitivity to extreme audience vote dispersion by discretizing score differences into bounded ordinal intervals. This mechanism mitigates structural amplification effects present in the percentage rule.

From an institutional design perspective, the shift toward ranking aggregation enhances robustness against popularity concentration while preserving audience participation. When combined with the Judges' Save safeguard, the mechanism corrects distributional distortions without eliminating viewer influence, thereby improving the alignment between competitive outcomes and professional evaluation.

Overall, the Season 27 case illustrates that percentage rules amplify demand-side concentration effects, whereas ranking rules provide a more robust institutional design by bounding the influence of audience preference shocks.

## 4. Dynamic Weight Allocation and Multi-Objective Mechanism Optimization

### 4.1. Dual-Objective Trade-off Model between Fairness and Excitement

To reconcile professional credibility with audience engagement, a dual-objective optimization framework is constructed to determine the optimal balance between judges' weight $w_J$ and audience weight $w_F$ [4] [5].

### 4.1.1. Objective Function Definitions

The Fairness Index $I_F \in [-1,1]$ is defined as the Spearman rank correlation coefficient between the final competition ranking and the cumulative judges' ranking. Formally,

$$I_F = \rho_{rank}(R_{final}, R_{judge}) \tag{7}$$

A value closer to 1 indicates stronger alignment with professional evaluation standards.

The Excitement Index captures audience engagement and is composed of two components:

Suspense: measured by the variance of weekly ranking changes;

Engagement: measured by the probability that contestants outside the top professional tier advance (denoted as upset probability).

The comprehensive entertainment index is defined as:

$$I_E = \alpha \cdot \mathrm{Var}(\Delta Rank) + \beta \cdot P(\mathrm{Upset}) \tag{8}$$

Where $\alpha$ and $\beta$ represent relative importance weights.

This formulation explicitly models the trade-off between professional consistency and demand-driven reversals.

### 4.1.2. Dynamic Weighting Mechanism

Rather than adopting a fixed weight ratio (e.g., 50/50), the model introduces a time-dependent weight function. Let $t \in [0,1]$ denote normalized season progress (0 at the start, 1 at the final) [6]. The total score is defined as:

$$S_{total}(t) = w_J(t) \cdot S_{Judge} + (1 - w_J(t)) \cdot S_{Fan} \tag{9}$$

The weight $w_J(t)$ follows a Sigmoid structure, gradually increasing over time. This design assigns greater audience influence in early stages—encouraging participation and visibility—while progressively strengthening professional authority in later rounds to ensure outcome legitimacy.

## 4.2. Pareto Frontier Analysis and Dynamic Mechanism Design

### 4.2.1. Pareto Frontier Analysis

Using historical data from Seasons 1–34, simulations are conducted under various $(w_J, w_F)$ combinations, generating corresponding $(I_F, I_E)$ outcome coordinates.

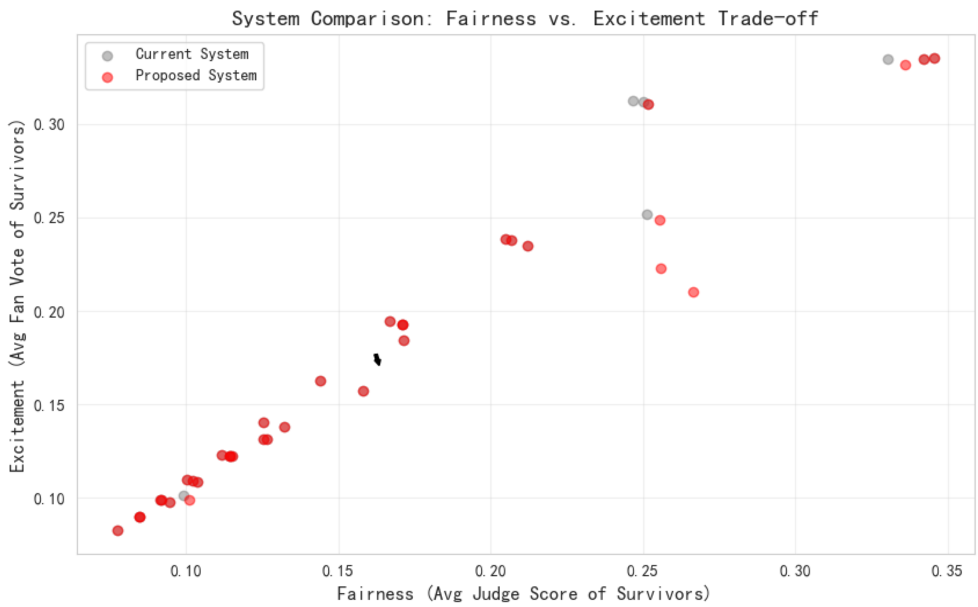


**Figure 3.** Pareto fairness vs excitement

From the Fig. 3, the resulting Pareto frontier reveals three distinct regimes:

High entertainment, low fairness: systems heavily dominated by audience voting, prone to popularity-driven reversals.

High fairness, low entertainment: systems dominated by professional scoring, characterized by limited volatility.

Intermediate frontier region: a balanced trade-off between professional integrity and audience-driven engagement.

Simulation results indicate that assigning judges' weight within the 60%–70% interval preserves more than 85% professional alignment while maintaining substantial ranking volatility. This intermediate zone represents an institutional equilibrium between credibility and commercial appeal.

### 4.2.2. Implementation Strategy: Dual-Track Dynamic Scheme

Based on optimization outcomes, a two-stage dynamic scoring structure is proposed:

Regular Season (Week 1–Semi-finals): a 50/50 weighted ranking system, preserving audience participation and allowing moderate reversals.

Final Stage: a 70/30 weighted ranking system combined with a Judges' Save safeguard, increasing professional dominance in championship determination.

Simulation results show that under this dynamic adjustment, early-stage volatility remains high, whereas late-stage championship probability for top professional performers increases from approximately 50% to above 90%.

This staged mechanism mitigates extreme popularity distortions without eliminating audience influence, as shown in the Fig. 4.

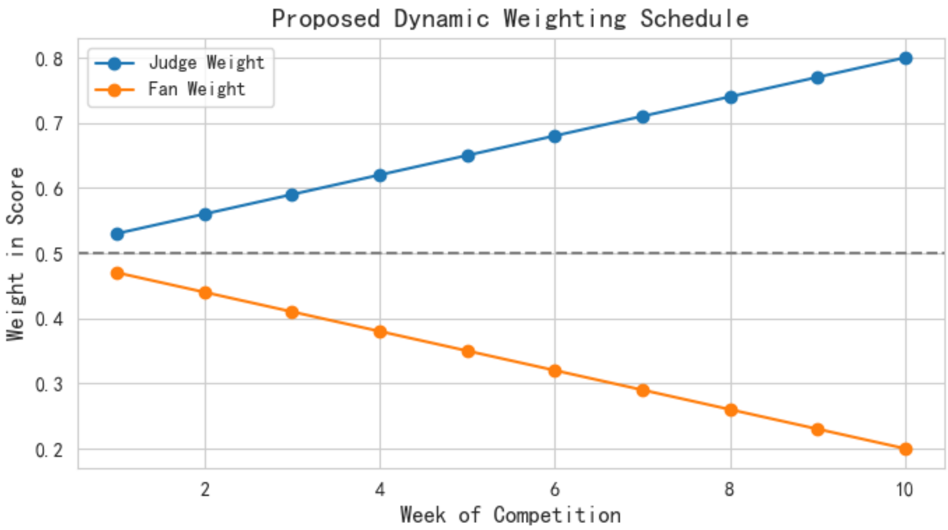


**Figure 4.** Dynamic weighting system

### 4.3. Institutional Implications of Dynamic Aggregation Mechanisms

The dynamic weighting framework resolves the structural tension between popularity and technical merit by aligning audience participation incentives with professional outcome stability across competition phases.

Gradual increases in professional weight enhance institutional credibility while preserving commercial engagement in earlier rounds. Such a mechanism design ensures that final outcomes reflect sustained professional quality rather than transient demand shocks, thereby strengthening both artistic standards and long-term audience trust.

## 5. Conclusion

This paper systematically analyzes the aggregation mechanism of multi-source evaluation information, revealing how different rule structures influence ranking stability and reversal probability. Results indicate that under the percentage summation mechanism, when preference distributions are highly concentrated, their contribution to fluctuations significantly exceeds that of expert scores, readily forming structural amplification. Conversely, the ranking mechanism effectively suppresses extreme shocks by discretizing data and confining discrepancies within fixed intervals. Simulation results demonstrate that setting the professional weight between 60% and 70% maintains over 85% consistency while preserving moderate volatility. Under the dynamic adjustment scheme, the winning probability of core participants in later stages increases from approximately 50% to over 90%, indicating that phased weight progression effectively balances stability and activity.

Overall, this paper provides a quantifiable analytical framework at the rule structure and weight configuration levels, offering empirical evidence for optimizing competitive mechanisms. Future research may further integrate richer behavioral data to test the model's external applicability.

## References

[1] Liu, W., Zhang, H., Li, Y., et al. Research on Answer Aggregation and Ranking in Socialized Q&A Communities Based on Semantic Networks [J]. Information Science, 2021, 39 (09): 94 - 100. DOI: 10.13833/j.issn.1007 - 7634. 2021.09.013.

[2] Wang Yuefen, Dong Xiaoyi, He Jin. Research on Generation and Evaluation of Scenario-Based Intelligence Service Schemes Based on Multi-Source Data Aggregation [J]. Research on Scientific and Technical Information, 2024, 6 (03): 56 - 68. DOI: 10.19809/j.cnki.kjqbyj. 2024.03.005.

[3] Jiang Guangyun. Monte Carlo-Based Intelligent Scheduling Model and Simulation for Medical Equipment [J]. Automation and Instrumentation, 2024, (12): 80 - 84. DOI: 10.14016/j.cnki.1001 - 9227.2024.12.080.

[4] Xia Tian, Shen Xinwei, Shang Yuwei. Multi-Objective Optimization Decision for Distribution Network Maintenance Based on Mixed-Integer Programming and Its Sensitivity Analysis [J]. Power System Technology, 2024, 48 (11): 4680 - 4689. DOI: 10.13335/j.1000 - 3673.pst. 2024.0076.

[5] Sun Yan. Intelligent Process Control Method Based on Dynamic Multi-Objective Optimization and Multi-Attribute Decision Making and Its Application [D]. Central South University, 2023. DOI: 10.27661/d.cnki.gzhnu.2023.005921.

[6] Wang Yuan, Zhang Zhihui, Ren Jie, et al. Comprehensive Safety Evaluation of Large Aqueducts Based on Dynamic Weight-Cloud Model [J]. Journal of Hohai University (Natural Science Edition), 2023, 51 (02): 56 - 63.