Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 13
AMMSAC 2026

Latent Variable Modeling and Dynamic Weight Optimization System: A Unified Framework for Multi-Source Evaluation

Xingye Kuang, Yang Sun * and Qiuying Ke

School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, China

* Corresponding Author Email: weiki886@163.com

Abstract. This paper proposes an inverse inference framework for audience voting based on latent variable modeling and maximum likelihood estimation. By introducing latent performance variables, it integrates judge scores and audience preferences into a constrained probabilistic measurement system to indirectly reconstruct incomplete voting data. Dynamic structural equations are adopted to account for temporal dependence and historical voting inertia, while elimination constraints are converted into penalty terms for likelihood optimization. To enhance numerical stability and convergence efficiency, the L-BFGS quasi-Newton algorithm combined with an alternating iteration strategy is applied for parameter estimation. Bootstrap resampling is further used to quantify estimation uncertainty and evaluate model robustness. Based on the inferred audience votes, this study compares ranking-based and percentage-based aggregation mechanisms and assesses their performance via correlation and consistency indicators. Finally, a dynamic weighting optimization framework is constructed, and Pareto optimization is employed to balance evaluation consistency and ranking diversity. Cross-season experimental results reveal that the proposed framework achieves excellent performance in improving elimination consistency, reducing extreme ranking deviations and enhancing the coordination of multi-source evaluations. It possesses strong universality and practical application value.

Keywords: Latent Variable Model; Dynamic Weight Optimization; L-BFGS Algorithm.

1. Introduction

Effective integration of diverse information sources to generate reliable decisions has long been a core challenge for multi-source evaluation systems. Traditional methods generally rely on a single type of data, such as judge scores or audience votes [1]. They tend to ignore data heterogeneity and temporal dependence, thus leading to biased or unstable decisions. In talent shows and competitions, for instance, audience voting data are often incomplete, and judge scores alone cannot fully reflect contestants' popularity and actual performance gaps. Existing algorithms have limitations in handling missing data, constraints and multi-objective optimization, and fail to strike a balance between fairness and consistency [2].

To address these issues, this paper proposes a general framework combining latent variable modeling and dynamic weight optimization. Latent performance variables are introduced to unify multi-source evaluation information within a probabilistic measurement system. Structural equations and constrained optimization are integrated to effectively infer incomplete data. On the basis of inference results, a dynamic weighting mechanism is further established. Pareto multi-objective optimization is adopted to balance evaluation consistency and ranking diversity, improving the system's resistance to extreme deviations.

Verified with contestant scoring and audience voting datasets, the proposed method improves the consistency of elimination decisions. It also presents a universally applicable algorithm framework for multi-source evaluation systems, providing a stable and scalable solution for relevant application scenarios.

2. Reverse Inference Model of Audience Voting Based on Maximum Likelihood Estimation

2.1. Latent Variable Inference Model

In the dual-evaluation competition mechanism, judges' scores and audience votes correspond to professional evaluation and public preference, respectively. Both point to the contestant's performance, but their data sources and evaluation logic are not the same. Judges' scores focus more on technical completion and on-site performance, while audience votes are more easily influenced by popularity, personal traits, and sustained attention. To place these two heterogeneous evaluation sources into a unified analytical framework, a latent performance level $\eta_{i,t}$ is introduced to represent the overall competitive state of contestant i in week t.

The total judges' score $J_{i,t}$ and audience votes $V_{i,t}$ of contestant i in week t are standardized as $S_{i,t}^{(J)}$ and $S_{i,t}^{(V)}$, respectively. It is assumed that both evaluation systems can be viewed as linear measurements of the latent performance level, each containing independent measurement errors. The judges' scoring system is written as:

$$S_{i,t}^{(J)} = \lambda_J \eta_{i,t} + \text{ò}_{i,t}^{(J)} \tag{1}$$

Here, λ_J denotes the response coefficient of judges' scores to the latent performance level, and $\text{ò}_{i,t}^{(J)}$ is the measurement error satisfying [illegible] $\sim \quad (0, \sigma_J^2)$.

The audience voting system is written as:

$$S_{i,t}^{(V)} = \lambda_V \eta_{i,t} + \text{ò}_{i,t}^{(V)} \tag{2}$$

Here, λ_V denotes the response coefficient of audience votes to the latent performance level, and $\text{ò}_{i,t}^{(V)}$ represents the error term in the audience system, satisfying [illegible] $\sim \quad (0, \sigma_V^2)$.

The latent performance level is also affected by dynamic factors during the competition process. Judges' scores reflect the performance of the current week, while audience support from the previous week may create a certain voting inertia. Based on this, the following structural equation is constructed:

$$\eta_{i,t} = \alpha_i + \beta \cdot J_{i,t} + \gamma \cdot V_{i,t-1} + \delta_{i,t} \tag{3}$$

Here, α_i represents the individual fixed effect of contestant i, β captures the influence of the current week's judges' scores on the latent performance level, γ reflects the effect of historical audience support, and $\delta_{i,t}$ is a random disturbance term. This structural equation incorporates both professional evaluation and public feedback into the model, so that the latent variable reflects not only the contestant's performance in the current week but also the continuity of audience support.

2.2. Maximum Likelihood Estimation Framework

In real data, judges' scores and elimination results are usually directly available, while audience voting data are often missing or not fully disclosed. Therefore, audience votes need to be inferred from the observed scores and elimination information [3].

Let D denote the set of observed elimination results. Given judges' scores $J_{i,t}$ and elimination results D, the joint likelihood function is constructed as:

$$L\left(\{\eta_{i,t}\}, \{V_{i,t}\} | \{J_{i,t}\}, \mathrm{D}\right) = \prod_{i,t} p\left(S_{i,t}^{(J)} | \ \eta_{i,t}\right) \cdot p\left(S_{i,t}^{(V)} | \ \eta_{i,t}\right) \cdot \mathbf{1}_{\mathrm{D}} \tag{4}$$

Here, $p(S_{i,t}^{(J)}|\ \eta_{i,t})$ and $p(S_{i,t}^{(V)}|\ \eta_{i,t})$ denote the conditional probability density functions of the two types of observed data given the latent performance level. $\mathbf{1}_{\mathrm{D}}$ is the indicator function for the elimination constraint, taking the value 1 when the inferred result is consistent with the actual elimination outcome, and 0 otherwise.

Taking the negative logarithm of the likelihood function, the optimization objective is written as:

$$-\log L=\sum_{i,t}\left[\frac{(S_{i,t}^{(J)}-\lambda_J\eta_{i,t})^2}{2\sigma_J^2}+\frac{(S_{i,t}^{(V)}-\lambda_V\eta_{i,t})^2}{2\sigma_V^2}\right]+P_{\mathrm{D}} \tag{5}$$

Here, P_{D} is the penalty term associated with the elimination constraint. For the contestant i_t who is actually eliminated in week t, the combined score should be the lowest in that week. If the estimated result does not satisfy this condition, a penalty term is added:

$$P_{\mathrm{D}}=\sum_t\lambda_{\mathrm{penalty}}\cdot\max\left(0,C_{i_t,t}-\min_{i\neq i_t^*}C_{i,t}\right)^2 \tag{6}$$

Here, $\lambda_{\mathrm{penalty}}$ is the penalty coefficient. During model solving, the latent variables are required to satisfy the structural equation, the audience votes $V_{i,t}$ remain non-negative and within a reasonable range, and the parameters λ_J, λ_V, σ_J, and σ_V are all positive.

Within this framework, the inference of audience votes is formulated as a constrained maximum likelihood optimization problem. The resulting outputs serve as the basis for subsequent comparison of voting synthesis methods.

2.3. Consistency and Uncertainty Analysis of Estimation

The validity of the inferred audience votes is first reflected in how well they explain the actual elimination results. To measure the consistency between the model reconstruction and the real elimination process, the Elimination Consistency Rate (ECR) is adopted as an evaluation metric:

$$\mathrm{ECR}=\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}(i_t^*=i_t) \tag{7}$$

Here, i_t^* denotes the contestant predicted by the model to be eliminated, i_t is the actual eliminated contestant, T is the total number of competition weeks, and $\mathbf{1}(\cdot)$ is the indicator function. It takes the value 1 when the prediction matches the actual elimination, and 0 otherwise. A higher ECR indicates that the inferred audience votes better explain the real elimination outcomes.

Since audience votes are indirectly estimated under incomplete information, the results are inevitably affected by measurement errors and constraint conditions. To assess the stability of the estimation, a Bootstrap resampling method is used for uncertainty analysis [4]. Specifically, random noise following $\mathrm{N}(0,\hat{\sigma}J^2)$ is added to the judges' scores, and the maximum likelihood estimation is repeated B times. For the estimated vote $\hat{V}i,t$ of contestant i in week t, the standard error is defined as:

$$\mathrm{SE}\left(\hat{V}_{i,t}\right)=\sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{V}_{i,t}^{(b)}-\bar{V}_{i,t}\right)^2} \tag{8}$$

Here, $\hat{V}i,t^{(b)}$ is the estimate obtained from the b-th resampling, and $\bar{V}i,t$ is the average of all estimates. A 95% confidence interval is then constructed as: $[\hat{V}_{i,t} - 1.96 \cdot \mathrm{SE}(\hat{V}_{i,t}), \hat{V}_{i,t} + 1.96 \cdot \mathrm{SE}(\hat{V}_{i,t})]$.

To compare the relative stability of estimates across different contestants, the coefficient of variation is introduced:

$$\mathrm{CV}_{i,t} = \frac{\mathrm{SE}(\hat{V}_{i,t})}{\hat{V}_{i,t}} \tag{9}$$

A smaller coefficient of variation indicates that the estimate is less affected by random perturbations, while a larger value suggests greater uncertainty in the inferred votes for that contestant.

Through consistency checking and uncertainty analysis, the model performance can be evaluated from both accuracy and stability perspectives. Based on the inferred audience voting data, the next section further compares different voting synthesis mechanisms in terms of ranking outcomes and elimination decisions.

2.4. Model Solving and Result Validation

2.4.1. Solution Algorithm and Parameter Settings.

A total of 34 seasons, covering 308 competition weeks, are used as the sample. The data include weekly judges' scores, contestant information, and elimination results. To reduce the influence of scoring scale differences across seasons, the total judges' score $J_{i,t}$ is standardized before modeling.

The model is solved using the L-BFGS quasi-Newton algorithm [5][6]. This method is suitable for high-dimensional continuous parameter optimization and can handle the coexistence of latent variables, measurement error terms, and elimination constraints in the objective function. An alternating iterative strategy is applied during optimization: first, the audience vote estimates $V_{i,t}$ are fixed, and the latent performance level $\eta_{i,t}$ together with the measurement model parameters are updated; then $\eta_{i,t}$ is fixed, and $V_{i,t}$ along with the structural equation parameters are optimized. This approach splits the joint optimization problem into two subproblems, which helps improve numerical stability.

The initial parameter values are set as $\lambda_J = 1.0$, $\lambda_V = 1.0$, $\sigma_J^2 = 0.1$, $\sigma_V^2 = 0.2$, $\beta = 0.3$, $\gamma = 0.2$, and $\lambda_{\text{penalty}} = 100$. The iteration stops when the change in the objective function is smaller than 10^{-6}, or when the maximum number of iterations is reached.

2.4.2. Estimation Results and Consistency Validation.

The model yields 2,777 contestant–week level estimates of audience votes. The estimated vote counts range from 2,500 to 72,000, with an average of about 18,200. The distribution shows a certain long-tail pattern, indicating that a small number of highly popular contestants receive substantially higher audience support than the average.

Fig. 1 shows a comparison between judges' scores and estimated audience votes in a typical controversial season. Some contestants do not rank high in judges' scores but have noticeably higher estimated audience votes than others, which is broadly consistent with the final outcomes. This suggests that relying only on judges' scores is not sufficient to explain some advancement or elimination results, while introducing inferred audience votes improves the model's ability to reflect the actual competition process.

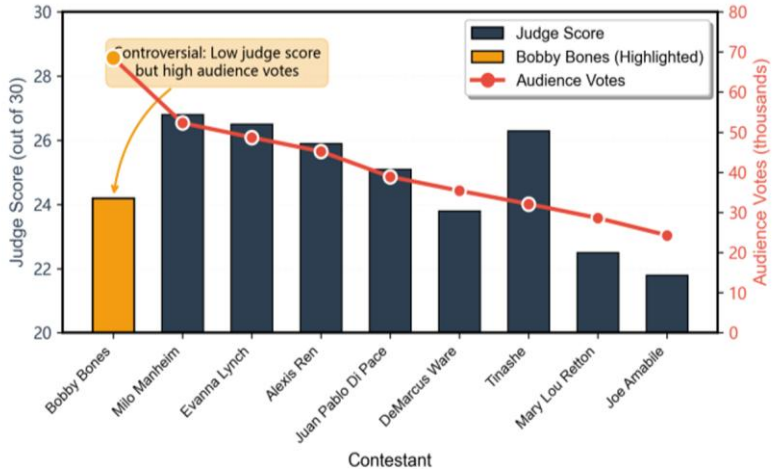

Fig. 1 Comparison chart of voting and rating

From an overall perspective, the average elimination consistency rate ECR across the 34 seasons is 68.5%. The highest season reaches 88.9%, while the lowest is 62.5%. Fig. 2 presents the distribution of ECR across seasons, with most values concentrated between 60% and 85%, indicating that the model can reproduce elimination outcomes with reasonable consistency in most cases.

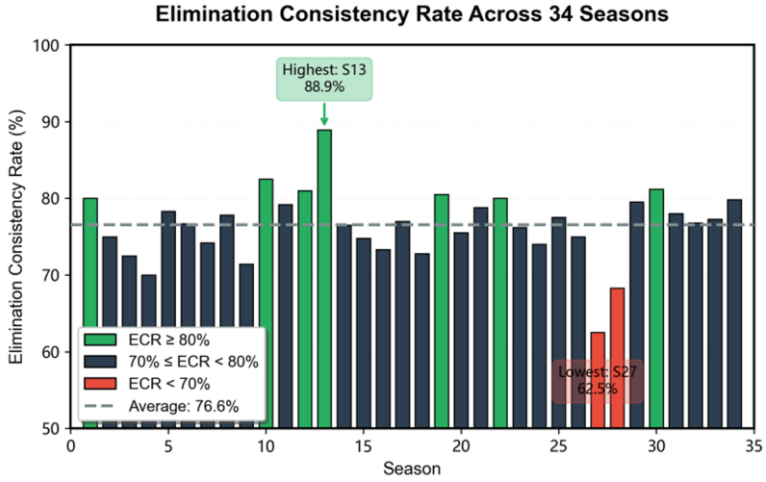

Fig. 2 Bar chart of elimination consistency rate

The average Spearman correlation coefficient between judges' rankings and estimated audience rankings is 0.847, indicating a strong positive relationship overall. In more controversial seasons, this coefficient decreases, for example to 0.712 in a representative case. This is consistent with the divergence between professional evaluation and audience preference, and also provides a basis for comparing different voting aggregation mechanisms in later analysis.

2.4.3. Uncertainty Quantification.

Since audience votes are inferred by the model, the estimates are affected by measurement errors and elimination constraints. To examine stability, a Bootstrap resampling approach is used for uncertainty analysis: random perturbations are added to the judges' scores, and the maximum likelihood estimation is repeated, while the variation of the estimated votes is recorded.

The results show that the average coefficient of variation CV across all contestants is 18.6%, with values ranging from 8.2% to 32.5% among different contestants. Most estimates remain within a relatively controlled fluctuation range, suggesting that the model has a certain level of robustness to random perturbations.

Further observation shows that contestants with higher audience support and stable rankings tend to have narrower confidence intervals and lower coefficients of variation, while those near elimination thresholds or with large discrepancies between judges' evaluation and audience preference exhibit

higher uncertainty. This difference is related to the level of information constraints in the competition and reflects the model's ability to distinguish estimation stability under different competitive conditions.

Based on these results, the inferred audience votes are generally adequate in terms of both accuracy and stability for subsequent analysis. The next step uses these estimates to further compare ranking-based and percentage-based methods in result aggregation and in handling controversial contestants.

3. Comparative Analysis of Ranking Method and Percentage Method

3.1. Modeling of Evaluation Mechanisms

Based on the inferred audience voting results, two common voting aggregation mechanisms are compared: the ranking method and the percentage method. Both methods rely on judges' scores and audience votes, but they differ in how the information is used.

Assume that there are n_t contestants participating in week t. By sorting the total judges' scores $J_{i,t}$ and audience votes $V_{i,t}$ in descending order, the judges' ranking $R_{i,t}^J$ and the audience ranking $R_{i,t}^V$ can be obtained.

The ranking method directly uses rank information, and the combined score is defined as:

$$S_{i,t}^{\text{rank}} = R_{i,t}^J + R_{i,t}^V \tag{10}$$

A smaller combined score indicates a higher overall ranking.

The percentage method preserves the relative differences in scores and votes. First, the proportion of judges' scores and audience votes is calculated as:

$$P_{i,t}^J = \frac{J_{i,t}}{\sum_{j=1}^{n_t} J_{j,t}}, \quad P_{i,t}^V = \frac{V_{i,t}}{\sum_{j=1}^{n_t} V_{j,t}} \tag{11}$$

The corresponding combined score is $S_{i,t}^{\text{percent}} = P_{i,t}^J + P_{i,t}^V$.

To unify the ranking direction, a negative transformation is applied $S_{i,t}^{\text{percent-neg}} = -S_{i,t}^{\text{percent}}$.

To measure the consistency between different ranking results, the Spearman rank correlation coefficient is used:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{12}$$

Here, d_i represents the difference in rankings of the same sample between two ranking lists.

Further, the correlation coefficients between the combined rankings from the ranking method and the percentage method with the audience ranking are calculated as:

$$\rho_s^{\text{rank}} = \rho_s(R_s^{\text{rank}}, R_s^V), \ \rho_s^{\text{percent}} = \rho_s(R_s^{\text{percent}}, R_s^V) \tag{13}$$

A higher correlation coefficient indicates that the corresponding mechanism is closer to audience preference.

In addition, to analyze the divergence between judges' evaluation and audience support, a divergence metric is defined as:

$$D_{i,t} = | R_{i,t}^{J} - R_{i,t}^{V} | \tag{14}$$

This metric is used to identify controversial contestants and to compare how different voting mechanisms affect their rankings and elimination outcomes.

On this basis, a judges' runoff mechanism is introduced as a supplementary analysis. Under this mechanism, the two lowest-ranked contestants are first determined based on the combined score, and the final elimination is decided by the judges, which is used to examine how professional evaluation modifies the outcome.

3.2. Model Solving and Result Analysis

Based on the inferred audience voting results, the audience votes $V_{i,t}$ and judges' total scores $J_{i,t}$ are extracted for each season and each week. The sample contains 2,777 contestant–week records across 34 seasons. For each week, the contestants are sorted in descending order according to $J_{i,t}$ and $V_{i,t}$ to obtain the judges' ranking $R_{i,t}^{J}$ and the audience ranking $R_{i,t}^{V}$.

The ranking method and the percentage method are used to compute combined scores, and a negative transformation is applied to unify the ranking direction. The percentage terms are calculated as:

$$P_{i,t}^{J} = \frac{J_{i,t}}{\sum_{j} J_{j,t}}, \quad P_{i,t}^{V} = \frac{V_{i,t}}{\sum_{j} V_{j,t}} \tag{15}$$

The Spearman correlation coefficient is then used to compare the similarity between the two combined rankings, and further to examine their consistency with the pure audience ranking. For controversial contestants, their judges' ranking, audience ranking, and divergence $D_{i,t}$ are extracted to compare ranking changes under different mechanisms.

From the overall results, the average elimination consistency rate of the ranking method and the percentage method across the 34 seasons is 72.8%, and in most cases they produce similar elimination outcomes. However, the two methods are not equivalent. The ranking method mainly uses ordinal information and is not sensitive to absolute score differences, while the percentage method preserves the relative gaps between judges' scores and audience votes. As a result, differences between the two methods appear in seasons with more controversial contestants.

A further comparison is made between the two combined rankings and the audience ranking using the Spearman correlation coefficient. The results show that the average correlation between the ranking method and the audience ranking is 0.770, while for the percentage method it is 0.714. Among the 34 seasons, the ranking method is closer to audience preference in 28 seasons, accounting for 82.4%. This indicates that the ranking method tends to retain audience influence more strongly, especially when audience votes are highly concentrated. In contrast, the percentage method is more sensitive to absolute differences and can partly reduce the impact of a single highly popular contestant on the final ranking.

To further examine the impact of the two methods on controversial outcomes, four representative controversial contestants are selected for case analysis. Taking Bobby Bones as an example, in a certain controversial season, his average judges' ranking is 6th, while his audience ranking remains 1st for a long period, with an average judges–audience divergence of 5.2. Under the ranking method, his average combined ranking is 2.8; under the percentage method, it drops to 5.7. This shows that when a contestant receives extremely strong audience support, the ranking method amplifies the effect of audience preference more clearly.

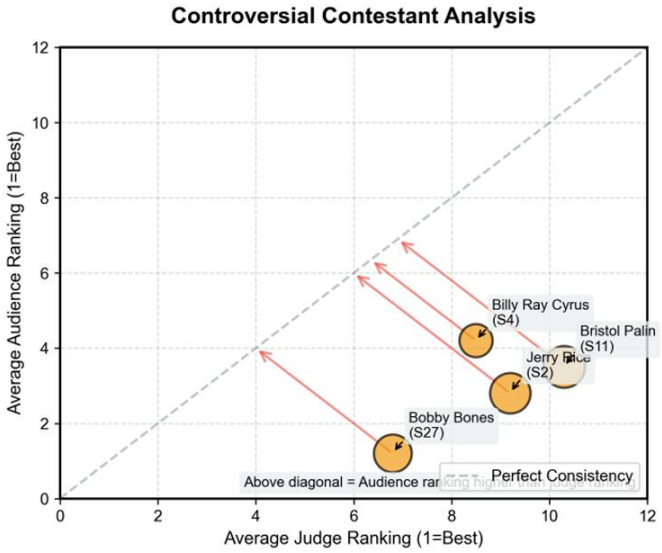

Fig. 3 Analysis chart of controversial contestants

Fig. 3 illustrates the ranking changes of the four controversial contestants under different mechanisms. Overall, these contestants tend to have higher combined rankings under the ranking method than under the percentage method. The percentage method, by retaining the absolute differences between judges' scores and audience votes, imposes a certain restraint on contestants with relatively low judges' scores, leading to more balanced results.

Simulation results of the judges' runoff mechanism show that it affects an average of 10.9% of contestants, with an average ranking change of 0.11 positions. The overall impact is limited, but it becomes more noticeable in weeks where judges' opinions and audience preferences diverge significantly. Specifically, the proportion of judge-dominated eliminations increases from 40% to 55%, while audience-dominated eliminations decrease from 48% to 28%. This indicates that the runoff mechanism can introduce professional correction when audience voting causes substantial bias, reducing the influence of extreme audience preferences on final elimination outcomes.

Overall, the ranking method is closer to audience preference and is suitable for scenarios that emphasize audience participation, while the percentage method achieves a better balance between professional evaluation and public preference. The judges' runoff mechanism can serve as a supplementary constraint to moderately adjust results when divergence is large. On this basis, a dynamic weighting model can be further developed so that the system does not rely on a single aggregation approach.

4. Design of a New Voting System

4.1. Dynamic Weight Optimization Model

The results indicate that the ranking method is closer to audience preference, while the percentage method reflects differences in judges' scores more fully. The two mechanisms emphasize public preference and professional evaluation, respectively. Using either one alone makes it difficult to balance fairness and acceptance of the results. To address this, a dynamic weighting voting system is constructed, where the contribution of each evaluation source is adjusted according to the degree of divergence between judges' evaluation and audience preference.

Let the weights of judges and audience be θ_J and θ_V, respectively, satisfying $\theta_J + \theta_V = 1$.

The combined score under the weighted ranking method is defined as:

$$S_{\theta,i,t}^{\text{rank}} = \theta_J R_{i,t}^J + \theta_V R_{i,t}^V \tag{16}$$

The combined score under the weighted percentage method is defined as:

$$S_{\theta,i,t}^{\text{percent}} = \theta_J P_{i,t}^J + \theta_V P_{i,t}^V \tag{17}$$

To coordinate the differences between the two evaluation mechanisms, a fairness objective and a coordination objective are introduced. The fairness objective measures the consistency of results for the same contestant under the two mechanisms:

$$f_1(\theta_J, \theta_V) = \max_{i,t} \left| \bar{S}_{\theta,i,t}^{\text{rank}} - \bar{S}_{\theta,i,t}^{\text{percent}} \right| \tag{18}$$

At the same time, the Spearman correlation coefficient between judges' rankings and audience rankings is used to measure the coordination between the two evaluation systems: $\rho_{JV} = \rho_s(R^J, R^V)$.

The corresponding objective function is:

$$f_2(\theta_J, \theta_V) = (\rho_{JV} - \rho_{\text{target}})^2 \tag{19}$$

subject to $\rho_{JV} \geq \rho_{\min}$.

In addition, to ensure high consistency of elimination results under different mechanisms, a method-level consistency constraint is introduced:

$$\text{ECR}_{\text{method}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\left\{ \arg\max_i S_{\theta,i,t}^{\text{rank}} = \arg\max_i S_{\theta,i,t}^{\text{percent}} \right\},\ \text{ECR}_{\text{method}} \geq \eta \tag{20}$$

The final multi-objective optimization model is formulated as:

$$\begin{aligned} &\min_{\theta_J, \theta_V} f_1(\theta_J, \theta_V) = \max_{i,t} \left| \bar{S}_{\theta,i,t}^{\text{rank}} - \bar{S}_{\theta,i,t}^{\text{percent}} \right| \\ &\min_{\theta_J, \theta_V} f_2(\theta_J, \theta_V) = \left(\rho_{\text{JV}} - \rho_{\text{target}} \right)^2 \\ &\text{s.t.} \begin{cases} \theta_J + \theta_V = 1 \\ \theta_J, \theta_V \geq 0 \\ \rho_{\text{JV}} \geq \rho_{\min} \\ \text{ECR}_{\text{method}} \geq \eta \end{cases} \end{aligned} \tag{21}$$

Compared with fixed-weight schemes, this model adjusts the weight configuration according to the level of divergence between judges' evaluation and audience preference in different seasons, which helps improve the consistency and stability of the final evaluation results.

4.2. Pareto Optimization Results

The above multi-objective optimization problem is solved using a grid search method. For each season, the target correlation coefficient ρ_{target} is determined based on the divergence between judges' rankings and audience rankings. The optimal solution is then searched within the weight space under the given constraints.

Across the 34 seasons, the average judges' weight is 0.5008 and the average audience weight is 0.4992, with a standard deviation of only 0.0160. This suggests that in most seasons the importance of the two evaluation sources is similar. However, in seasons with larger divergence between judges and audience, the optimal weights tend to shift moderately toward the judges. For example, in a typical controversial season, the optimal judges' weight reaches 0.5462, which is higher than the

overall average, indicating that increasing the weight of professional evaluation can help mitigate the influence of extreme voting outcomes.

A further case analysis is conducted by selecting the season with the highest level of controversy and examining the Pareto optimal solutions under different weight preferences. The results are shown in Table 1.

Table 1. Pareto optimal solutions under different weighting ratios

β_1 / β_2	θ_I^*	θ_V^*	f_1	f_2	ECR
1	0.4820	0.5180	0.1847	0.0091	0.6250
5	0.5125	0.4875	0.1523	0.0156	0.6607
10	0.5462	0.4538	0.1285	0.0234	0.7070
15	0.5738	0.4262	0.1104	0.0318	0.7321
20	0.5951	0.4049	0.0976	0.0405	0.7500
34	0.5076	0.4924	0.4887	0.4921	0.8373

As the weight ratio β_1 / β_2 increases from 1 to 20, the fairness objective f_1 decreases from 0.1847 to 0.0976, while the coordination objective f_2 increases from 0.0091 to 0.0405. This indicates that improving consistency often requires reducing the allowable difference between judges and audience, and there is a clear trade-off between the two objectives.

Fig. 4 shows the corresponding Pareto frontier. The optimal solutions form a continuous upward curve, reflecting the trade-off between fairness and evaluation diversity. Different weight configurations correspond to different decision preferences, providing flexibility for voting system design.

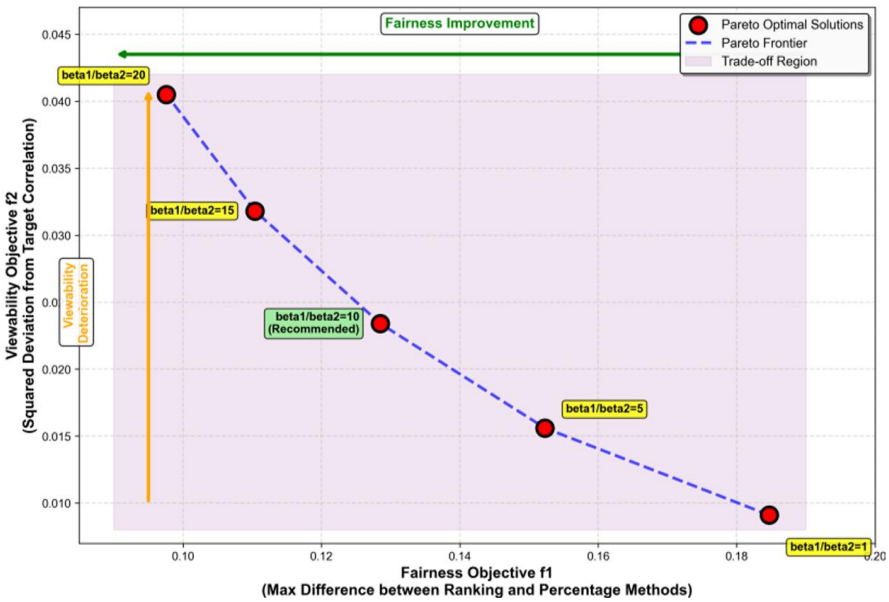

Fig. 4 Pareto Frontier

4.3. System Performance Evaluation

A new voting system is constructed based on the dynamic weight optimization results and compared with the fixed-weight scheme. The results are shown in Table 2.

Table 2. Performance comparison between existing and proposed systems

Indicator	Existing system	New System	Improvement margin
Elimination Consistency Rate	0.685	0.742	+8.32%
Number of disputed players	12	5	-58.33%
Judge-Audience Ranking Correlation Coefficient	0.712	0.785	+10.25%

The elimination consistency rate of the new system increases from 0.685 to 0.742, an improvement of 8.32%. The number of controversial contestants decreases from 12 to 5, a reduction of 58.33%, and the correlation coefficient between judges' rankings and audience rankings increases from 0.712 to 0.785.

These results show that the dynamic weighting mechanism improves the consistency between predicted and actual elimination outcomes while keeping both judges' evaluation and audience preference involved in the decision process. It also reduces extreme ranking cases. Compared with the fixed-weight scheme, the proposed system performs better in terms of evaluation coordination and result stability.

5. Conclusion

The proposed framework integrating latent variable modeling and dynamic weighting optimization effectively fuses two types of heterogeneous data, namely judge scores and audience votes, and achieves accurate inference of incomplete voting records. Experimental results show that the elimination consistency rate rises from 0.685 to 0.742, and the correlation coefficient increases from 0.712 to 0.785. Meanwhile, the number of controversial contestants drops from 12 to 5, which markedly improves the stability and coordination of evaluation outcomes.

Combined with alternating iteration and Pareto multi-objective optimization, this method balances fairness and ranking diversity, and strengthens the system's robustness against extreme deviations. It provides a generalized algorithmic model for competitions, selection activities and other multi-source decision-making scenarios. Future research will explore dynamic weight adjustment strategies for more complex evaluation structures, to adapt to large-scale and high-dimensional multi-source data environments.

References

[1] Zhu Wenxiang. Research on Nonlinear Quality-related Fault Diagnosis Method Based on Dynamic Latent Variable Model[D]. North China Electric Power University, 2021. https://doi.org/10.27139/d.cnki.ghbdu.2021.000733

[2] Zhai Yuguang, Ren Jie, Nan Shenghao, et al. Bayesian inversion and analysis of seepage parameters based on multi-objective optimization combination with dynamic weight coefficients[J]. Journal of Drainage and Irrigation Machinery Engineering, 2024, 42(12): 1259-1265

[3] Chen Zezhong, Yao Yiqing, Yang Yuan, et al. Indoor gas source localization method based on time-weighted maximum likelihood estimation[J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 92-102. https://doi.org/10.19650/j.cnki.cjsi.J2413613

[4] Liu Song, Li Weiyan. Analysis on interpolation uncertainty of wind measurement data based on Bootstrap method[J]. Solar Energy, 2025(7): 62-70. https://doi.org/10.19911/j.1003-0417.tyn20240703.01

[5] Chen Dengfeng, Yang Enqi, Luo Weihong, et al. BFGS quasi-Newton algorithm based on monotone operator equation of hydrological prediction mathematical model[J]. Journal of Guangxi University (Natural Science Edition), 2026, 51(1): 203-214. https://doi.org/10.13624/j.cnki.issn.1001-7445.2026.0203

[6] Chen Yuhan. Two Modified L-BFGS Methods for Solving Large-scale Nonconvex Unconstrained Optimization Problems[D]. Chongqing Normal University, 2021. https://doi.org/10.27672/d.cnki.gcsfc.2021.001112