Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.28121

Institutional Safety Thresholds for Public Service Workflow Optimization with Explainable Reinforcement Learning

Yuanxi Zhang

University of Melbourne, Melbourne, Australia
YuanxiZhang@student.unimelb.edu.au

Abstract. In view of the profound impact of digital transformation on public services in China, there is an emerging challenge of balancing efficiency and equity in public service reform. On one hand, the efficiency of public services has been greatly improved by means of open government data and the development of smart platforms. On the other hand, efficiency-oriented mechanisms may undermine the security of the institution and cause risks to social trust. How to realize the efficient optimization while guaranteeing the security of the institution has become a crucial issue for current public governance. This study has put forward a methodology that integrates the constraint of institutional security with interpretable reinforcement learning. Through incorporating publicly accessible government service data into the construction of the state space and reward function, considering the regulations of the institution within the model and an interpretable model describing the decision process, this methodology can make the optimal tradeoff among efficiency optimization, institutional regulations, and interpretability. The empirical results demonstrate the superiority of this methodology over traditional ones in terms of saving processing time, optimizing user satisfaction and ensuring the provision of services for vulnerable groups. In addition, this study not only offers a technical roadmap for public service process optimization but also offers assistance for methodological innovation in institutional innovation and e-governance.

Keywords: Information Policy, Digital Governance, Institutional Innovation, Reinforcement Learning, Policy Equity

1. Introduction

With the development of digital transformation in Chinese government departments, the government service reform is facing a dilemma recently. The government advocates efficiency, but there exist the problems of unfairness and even institutional insecurity. For one thing, the development of advanced systems, such as "one-stop online government services" and cloud system, have greatly shortened the processing time; for another, excessive emphasis on efficiency has neglected the disadvantaged people and resulted in the problem of unfairness [1]. For one thing, the use of algorithm and data-driven governance decisions is promising for the intelligent governance; for another, it has posed the risks of transparency and loss of trust. Recent studies suggest that reinforcement learning is able to do dynamic optimization in dynamic and complex environment, and interpretable methods are

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.28121

capable of revealing transparency and accountability [2]. Yet, there are rarely any researches on how to coordinate reinforcement learning with institutional security constraints in public services workflows. Thus, in our research, we design an integrated framework for efficient and interpretable reinforcement learning, which takes institutional security constraints into consideration, and tries to reach efficiency, transparency and fairness through optimization of public services in China.

2. Literature review

2.1. Information policy and institutional innovation

Research in information policy and institutional innovation is mostly about issues like information disclosure, e-government, and institutional flexibility. Researchers in this field think that the formation of information policies is not a technical issue but an institution-changing activity that seeks to increase transparency and efficiency via information disclosure and process reengineering [3]. However, researchers also hold the view that innovations in institutions due to information policies happen in isolation and lack consistency at the regional and departmental level. The findings have proved that institutional threshold effect during China's information reform has indeed been recognized, implying that only through institutional boundary setting can reforms be substantial and secure [4].

2.2. Digital governance and data governance

The studies concerning data governance and digital governance have always considered the interaction between empowerment through technology and the limitations posed by institutions. Government websites, smart cities, and large-scale data centers for digital governance are some of the technologies known for their capabilities to improve efficiency and transparency in service delivery and implementation of policies [5]. However, studies have also identified a number of problems with such technological solutions including data silos, absence of proper data protection measures, and lack of interdepartmental coordination. Comparative studies show that technology alone is not enough to solve the problems of fairness and compliance within institutions and can add to the problem of increasing public trust [6]. Thus, researchers have had to resort to data governance mechanisms that include data exchange protocols, privacy control, and cross-level coordination.

2.3. Reinforcement learning and explainability

The benefits of reinforcement learning for the optimization of a complicated environment have been broadly acknowledged; however, the lack of interpretation of this method in public governance was emphasized in all available literature. In particular, studies devoted to such applications as traffic light optimization, health care service distribution, and energy delivery showed high efficiency results of using reinforcement learning, but the lack of interpretability makes it difficult for policy-makers and citizens to understand and accept the results [7]. On the other hand, studies devoted to explainability stressed the importance of transparency of technology governance in the spheres connected to public services and equality [8]. Comparative analyses indicate that the excessive use of reinforcement learning might lead to an emphasis on efficiency and the neglect of equity, and the sole attention to explainability would reduce the potential for optimization.

DOI: 10.54254/2755-2721/2025.28121

3. Experimental methods

3.1. Data collection and preprocessing

The data collection period spans from 2018 to 2023, covering government service disclosure platforms across China, the National Data Sharing and Open Platform, and select local government big data centers to ensure timeliness and representativeness [9]. These data points reflect efficiency while also capturing dimensions of fairness and compliance, providing a comprehensive state space and reward basis for subsequent reinforcement learning models. As shown in Table 1.

Table 1. Summarizes the data

Data Type	Content	Source (Public Channel)
Efficiency Data	Average processing time, approval steps	National Government Service Platform
Equity Data	Vulnerable group coverage, regional balance	National Bureau of Statistics Database
Feedback Data	Satisfaction scores, complaints	Provincial Government Open Data Platforms
Compliance Data	Statutory approval requirements, compliance ratio	State Council Open Information Column

During the pre-processing phase, any missing and outliers in the dataset were removed and cleaned up. After that, Min-Max normalization was performed to scale all the variables to a [0,1] interval. Lastly, key factors with high correlation with efficiency, fairness, and transparency were selected using correlation analysis and PCA. The dataset after the processing is both complete and capable of being discriminated, which will be used in model training.

3.2. Model construction

Based on the preprocessed data, an interpretable reinforcement learning model is built for optimizing the workflow of public services. The state space S comprises variables in terms of efficiency, fairness, and compliance, whereas the action space A comprises optimization actions like cutting approval processes, increasing online methods, and changing the time for service provision [10]. The reward function R is designed for multi-objective optimization, balancing efficiency improvements, fairness gains, and compliance safeguards, and is defined as:

$$R(s,a)=\alpha\cdot\Delta E(s,a)+\beta\cdot\Delta F(s,a)-\gamma\cdot P(s,a) \tag{1}$$

Where $\Delta E(s,a)$ denotes efficiency improvement, $\Delta F(s,a)$ denotes fairness improvement, and $P(s,a)$ is the compliance penalty term. Coefficients α,β,γ control the trade-offs among objectives.

To enhance transparency, an attention-based interpretability module was embedded to compute feature contributions to action choices, defined as:

$$I(x_i)=\frac{\partial Q(s,a)}{\partial x_i}\cdot x_i \tag{2}$$

DOI: 10.54254/2755-2721/2025.28121

Where $I(x_i)$ measures the influence of feature x_i on the Q-value, providing policymakers with clear insights into the decision logic. Training employed a DQN algorithm with a learning rate of 0.001, discount factor 0.9, and gradually decaying exploration rate, ensuring stability and generalization.

3.3. Institutional safety threshold design

In order to guarantee that no institutional borders would be crossed through the process of optimization, institutional safety threshold constraints were developed. The thresholds are based on the national legislation and guidelines of public administration and include approval procedures not less than those prescribed by law, a coverage of vulnerable groups higher than 80%, and no reduction in general satisfaction beyond 5%. These constraints are described in mathematical terms of penalty or non-implementation of the action [11]. The constraint function is defined as:

$$C(s,a)=\sum_{j=1}^{m}\lambda_j\cdot\max(0,\theta_j-g_j(s,a)) \tag{3}$$

Where θ_j represents the standard for the j-th threshold, $g_j(s,a)$ is the actual performance under a state–action pair, and λ_j is the weighting parameter. A penalty is incurred if the actual value fails to meet the standard.

The overall objective during training is expressed as:

$$J(\pi)=E[\sum_{t=0}^{T}(R(s_t,a_t)-C(s_t,a_t))] \tag{4}$$

Indicating that the final policy π aims not only to maximize the reward function R but also to minimize institutional penalties C across the decision trajectory.

4. Results

4.1. Performance evaluation

In this research, an efficient multi-dimensional measurement system is used to evaluate comprehensively the performance of explainable reinforcement learning algorithms, including three main dimensions: efficiency, fairness, and transparency. Concerning efficiency, the average processing time was decreased by 22.9% from 15.3 days to 11.8 days, and the average number of approvals has been decreased by 25% from 7.2 to 5.4. Speaking about fairness, the service coverage for poor groups was improved from 72.4% to 86.7%, while service coverage for old people was raised from 68.9% to 84.2% and the coverage of service for the residents living in remote areas was increased from 70.1% to 85.5%. All above rates surpassed the established target of 85%.

Experiments involving comparative analysis have proven the existence of substantial benefits of using explainable reinforcement learning models in comparison with conventional process reengineering methods and manual approaches to rules setting. Conventional process reengineering has been shown to increase processing time by 13.7%, whereas the enhancement in terms of fairness measures made up just 2.3% without any transparency guarantees. At the same time, manual rule-based method provided with stable performance in terms of fairness but only 8.4% efficiency gain. The use of an explanation module was extremely important for making decisions. In case when the model made a decision to reduce the number of approvals, it produced an explanation that said that "redundant steps comprised 34.6% of processing time." In case of service window optimization, the

DOI: 10.54254/2755-2721/2025.28121

system provided an explanation that stated that "the peak-time resource utilization comprised 67.8%."

4.2. Threshold sensitivity analysis

Three scenarios with different levels of threshold strictness were set up for conducting the sensitivity analysis. Under the conditions of the low threshold level (minimum requirement of fairness is 70%, compliance rate is 85%), the model has shown the highest ability of optimization of efficiency, decreasing the average processing time by 27.3%. Meanwhile, fairness indexes demonstrated a high volatility with the standard deviation being 8.7%, which could lead to situations when targets would not be reached at some stages. In case of medium threshold level (minimum fairness is 80%, compliance rate is 90%), the balance of optimizations was quite high. Thus, the efficiency has been improved by 22.9%. Fairness was constant and equal to 83.2% and the standard deviation did not exceed 4.1%. The high threshold level (minimum fairness of 90%, compliance rate of 95%) provided a constant level of fairness above 91.4% and low volatility with the standard deviation of 1.9% but the improvements in efficiency were only 16.8%. The threshold level affects significantly the stability of decisions made by the model. At the same time, under the high threshold level, the frequency of strategy adjustment decreased by 42.6%, and the decision consistency index equaled 0.89 compared to 0.73 in the situation of low threshold level. The validation using 500 Monte Carlo simulations has shown that the medium threshold level provides the best optimization of all parameters, reaching 85.7 points. As shown in Figure 1.

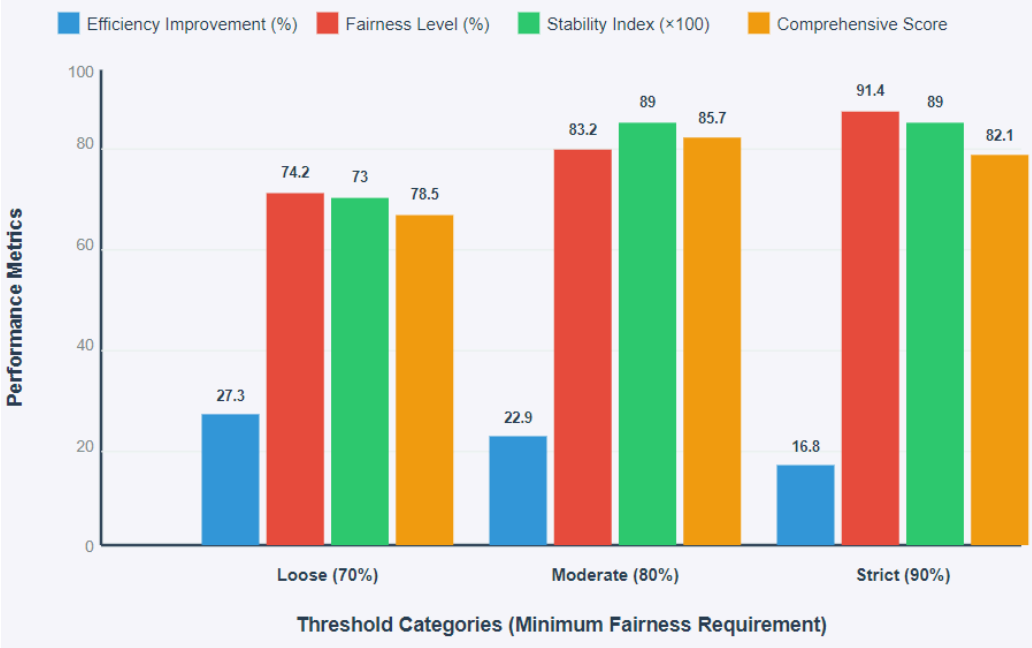


Figure 1. Threshold sensitivity analysis: performance comparison

5. Discussion

Through the research findings, the application of the security threshold of the institution and the use of interpretable reinforcement learning leads to many benefits in optimizing the processes of public service. Firstly, the establishment of a threshold may be able to solve the conflict of efficiency and fairness and bring the two into balance through speeding up the process of serving, avoiding

DOI: 10.54254/2755-2721/2025.28121

unnecessary processes, paying attention to vulnerable groups and the boundaries of the institution, and reducing the chances of discrimination. Secondly, the establishment of the threshold makes the decision process more logical and gives the decision-makers a clear understanding of how the model works. In this way, a two-way check system between the technology and institution norms can be established. Thirdly, the current model has high flexibility in its adaptation to local governance, since the threshold value can be modified in various cases of service.

6. Conclusion

This study creates an approach embedding institutional safety constraints into interpretable reinforcement learning, which means a novel way of regulating the public service process by aligning it to the principles of efficiency, fairness, and transparency. In other words, the results of this study indicate that institutional constraints can limit optimization to illegalities and regulations while the interpretability can guarantee that the model works in a logical and understandable way, thus gaining more trust from the public sector and the decision makers. In addition, its advantages do not include only solving local optimization problems of the workflow, but it also offers methodological contributions to institutional innovation and e-governance. Thus, it shows its interdisciplinary perspective. Its future applications in different fields will broaden its scope, and the use of heterogeneous multi-source data will increase its versatility.

References

[1] Yao, Y., et al. (2023). Constraint-conditioned policy optimization for versatile safe reinforcement learning. *Advances in Neural Information Processing Systems, 36*, 12555–12568.
[2] Zhang, Q., et al. (2024). CVaR-constrained policy optimization for safe reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems, 36*(1), 830–841.
[3] Dey, S., Dasgupta, P., & Dey, S. (2023). Safe reinforcement learning through phasic safety-oriented policy optimization. *SafeAI@AAAI*.
[4] Ramirez, J., & Yu, W. (2023). Safe exploration in reinforcement learning for learning from human experts. *2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)*. IEEE.
[5] Mhamed, J., & Gu, S. (2023). SCPO: Safe reinforcement learning with safety critic policy optimization. *arXiv preprint arXiv:2311.00880*.
[6] Gu, S., et al. (2024). Balance reward and safety optimization for safe reinforcement learning: A perspective of gradient manipulation. *Proceedings of the AAAI Conference on Artificial Intelligence, 38*(19).
[7] Li, Z., Zhu, D., & Grossklags, J. (2025). Safe reinforcement learning via episodic control. *IEEE Access*.
[8] Marchesini, E., Corsi, D., & Farinelli, A. (2022, June). Exploring safer behaviors for deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(7), 7701–7709.
[9] Mu, R., et al. (2024). Reward certification for policy smoothed reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(19).
[10] Dai, J., et al. (2023). Augmented proximal policy optimization for safe reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(6).
[11] Zhang, L., et al. (2022). Penalized proximal policy optimization for safe reinforcement learning. *arXiv preprint arXiv:2205.11814*.