

Adaptive Multi-User Preference Aggregation Using MORL and Neural User Models for Intelligent Appliance Control

Kalyanaraman P^{1}, Sadiya Khan¹, Yashas BN¹, Pramit Ghosh¹, and Siddesh Kedar¹*

¹School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

Abstract. The emergence of intelligent models has improved user preference modelling for personalized automation. Nevertheless, energy optimization frameworks developed using multi-objective reinforcement learning (MORL) have been designed and utilized mostly for single-user problems, neglecting the increasingly multi-user nature of the real-world setting (e.g., smart homes). This proposal presents a multi-user preference aggregation system based on the MORL framework developed to aggregate, equilibrate, and incorporate divergent specifications of multiple users on appliance performance into user agreements in the smart home. Each user is represented by a model, either an artificial neural network (ANN) or a convolutional neural network (CNN), which has been constructed and trained to predict a vector output representing the user's preferences at the appliance-level, e.g., temperature, time to wash, cooling intensity. These user-related vectors of multiple users (possibly all) can then be aggregated using a central MORL model, which will learn to combine and cluster user preferences in an optimal manner, dynamically. The MORL agent will receive reward signals based on user responses to appliances; positive responses will reinforce the aggregation policy, and negative responses will incur retraining of the associated user model. The two-way feedback model represents how the preference agent learns through the environment and the user disagreement to obtain the personalized equilibrium across all dimensions of preference from multiple users (possibly all) attributes. Instead of using NLP to adapt to preferences in text-based works, the suggested method emphasizes reinforcement-driven meta-learning for the purpose of adaptive preference fusion. It allows for scalable personalization where group dynamics and conflicts can be resolved automatically. The anticipated benefits include higher collective satisfaction, faster convergence of user-specific models, and more adaptability of the system over time. By reformulating user-centred optimization as a multi-agent coordination problem, this study expands the focus of MORL beyond energy efficiency toward a general, preference-aware intelligent control architecture.

1. Introduction

While intelligent automation is being increasingly integrated into daily life, human preference-machine control interaction has become an important design focus. Smart environments such as homes, workplaces, and public infrastructures today use adaptive control systems for regulating everything from energy to comfort. Yet, most of the state-of-the-art reinforcement learning-based frameworks are focused on single-user optimization. In reality, environments are inherently multi-user, having overlapping, dynamic, and sometimes conflicting preferences.

*Corresponding Author: pkalyanaraman@vit.ac.in

Recent studies have shown that MORL coupled with NLP can personalize energy optimization by user-specific constraints [8]. Their system fine-tunes energy usage patterns by incorporating linguistic feedback into a reward function. However, the system assumes that all user commands reflect a unified objective. In shared living environments, this assumption rarely holds.

Consider five inhabitants, each with unique preferences over air conditioning, the settings of a refrigerator, or the timing of laundry. A static MORL agent cannot optimally satisfy all of them [3] without any provision for learning from distributed feedback. The system must aggregate multiple neural representations of preference and continuously balance the satisfaction across users.

It upgrades the MORL framework to a multi-agent, user-adaptive architecture of learning agents. Every user is represented by a personalized neural network, either an ANN or a CNN, trained with interaction data to output a preference vector.

These vectors form the state space input to a central MORL agent, which learns to generate an aggregated control action vector. Most importantly, the MORL agent receives feedback through user satisfaction signals. When users collectively approve, the system reinforces its policy. In case a subset of users expresses dissatisfaction, the MORL agent identifies which of the user vectors contributed most to the rejected outcome and selectively retrains those models in an iterative process that allows self-organizing personalization to emerge. In that sense, it would blend ensemble learning principles with reinforcement-driven coordination.

This system overcomes three major challenges in existing MORL designs by introducing user-level retraining and the ability to propagate feedback.

1. Models lack inter-user differentiation in input.
2. Lack of focused adaptation if dissatisfaction occurs.
3. Static weighting of user influence in decision-making.

The work hereby offers a vision for personalization from single users to multi-user collectives that achieve consensus, fairness, and learning stability all at once [2]. The proposed approach generalizes MORL into a collaborative, feedback-driven framework capable of managing dynamic human-machine ecosystems.

2. Literature Review

Earlier studies on energy optimization and user-centred MORL have been about single-user systems. Reinforcement learning has previously been used in some studies for building or appliance-level control [1, 7, 9]. Liu et al. (2020) used Double Deep Q-Networks (DDQN) to optimize home energy costs with an up to 59% reduction in costs. Chen et al. (2021) examined incorporating user preferences in energy demand response ideas utilizing MORL, achieving your typical comfort-to-cost trade-offs for actions [7, 8]. However, both studies used paradigms on single-agent reinforcement and did not feature approaches to incorporating multi-user preference diversity. Xu et al. (2022) and Anvari-Moghaddam et al. (2014) also investigated MORL for microgrid optimization [4, 20]. Although there was an opportunity for optimizing between efficiency and comfort, there was no aspect of human-level personalization. Meanwhile, effective ensemble learning frameworks have demonstrated how multiple models could be explored in parallel or series of processes, to improve prediction accuracy and stability (Dietterich, 2000) [12, 18]. These methods are static aggregators for different models and involve no dynamic adaptation based on

reinforcement feedback. They do not take into account cross models dependencies, or interactive signals of satisfaction. These limitations in methods have created a shift toward more reinforcement based meta-ensembles, whereby, the model(s) have an experienced learning-for-combination process to the prediction, rather than a fixed weighted approach.

New developments in MARL can provide a framework for understanding multi-users optimization. Timilsina et al. (2021) developed an energy-sharing system based on reinforcement learning, which incorporated user preferences and thus improved efficiency in distributed energy trading [13]. Zhou et al. (2016) developed negotiation- based knowledge transfer among agents as a means of reducing the cost of coordination on sparse states, when using the shared representations of the knowledge of other agents [16]. These developments suggest a pathway to collaborative learning, in which many agents (users or user models) learn from interacting with a meta-agent, all with the shared goal of maximizing a communal reward. [13, 16]

Building on the idea of personalization, Gupta et al. (2020, 2022) demonstrated that the trade-off between saving energy, and user comfort is possible using user-centric MORL actually modelling user preferences [14]. The proposed architecture will build on these papers in the following ways:

1. It will reframe users as learning agents (through individual CNN/ANN models) instead of static inputs.
2. It will modularize the MORL policy applied at a metalevel that learns to combine and group user preferences.
3. It will design a feedback system in which dissatisfaction leads to agent retraining based on user preference, introducing a self-correction mechanism at both the global (meta-policy) and local (users) level.

In this way, the MORL system evolves from a single-agent optimizer towards the notion of a distributed learning, predictive system, aligned on the principles of ensemble and meta-learning.

3. Methodology

The proposed system architecture consists of three primary layers: User Model Layer, MORL Aggregator Layer, and Feedback Adaptation Layer.

Dataset Description

The research used a synthetically created dataset which contained multiple user preference information for its experimental assessment [6, 10]. The system assigned every user a latent preference vector which exists within a d -dimensional space that controls appliance operation through parameters like temperature and power level and operational duration. The researchers created training samples by applying controlled Gaussian noise to the latent vectors which produced realistic behaviour patterns. The researchers created 1000 preference samples for each of the $N = 5$ users and divided the samples into training and testing sets according to a 90:10 ratio. The synthetic dataset enables controlled evaluation of convergence, fairness, and consensus behaviour while preserving statistical consistency across users. The experimental setup permits researchers to conduct detailed studies on the MORL aggregation system without any environmental disturbances.

User Model Layer

Each user is represented by an independent neural network (CNN or ANN) trained to output a preference vector $V_i \in \mathbb{R}^d$, summarizing their preferences for multiple appliances (e.g., AC temperature, refrigerator power, washing cycle length) [15, 17]. These models are trained on historical user data or interactions and continuously fine-tuned based on MORL feedback.

MORL Aggregator Layer

The central MORL agent acts as a meta-policy network π_θ , where N is the number of users. It receives the concatenated or pooled user vectors and outputs an aggregation weight distribution $W = [w_1, w_2, \dots, w_N]$. The final action vector is computed as a weighted combination:

$$V_{agg} = \sum_{i=1}^N w_i V_i \quad (1)$$

The MORL policy is trained via reinforcement learning (e.g., REINFORCE or PPO) using collective user satisfaction as the reward [11, 19]. The reward signal incorporates both group satisfaction and individual fairness:

$$R_t = \lambda \bar{s} + (1 - \lambda) \text{Fairness} \quad (2)$$

where s denotes the average user satisfaction score and λ balances collective utility and fairness.

Feedback Adaptation Layer

Upon presenting the aggregated vector to users, binary feedback (satisfied/unsatisfied) or scalar ratings are collected.

- If satisfied: The MORL model is rewarded, reinforcing its current weighting scheme.
- If unsatisfied: The MORL model receives a penalty and computes contribution gradients to identify the user model(s) most responsible for dissatisfaction. Those specific user models are retrained locally.

Retraining focuses only on the most influential user:

$$i^* = \arg \max_i w_i \quad (3)$$

This ensures targeted refinement rather than global destabilization.

Clustering Mechanism

Over time, the MORL agent implicitly clusters users with similar preference vectors by learning stable weight distributions. Users with similar embeddings tend to receive correlated weights, resulting in emergent grouping behaviour.

Training and Optimization

Policy parameters θ are updated using policy gradient optimization [12, 18]:

$$\nabla_{\theta} J(\theta) = \mathbb{E} [\nabla_{\theta} \log \pi_{\theta}(a|S)(R_t - b)] \tag{4}$$

where b is a moving baseline to reduce gradient variance.

User models are optimized asynchronously using smaller learning rates to ensure stability.

Evaluation Metrics

System performance is measured through:

- Average group satisfaction score
- Convergence rate of user models
- Fairness (variance of satisfaction across users)
- Computational efficiency of MORL policy updates

Mathematical Interpretation of Consensus Formation

To formalize satisfaction, cosine similarity is used to measure alignment between the aggregated vector and each user preference:

$$\text{sim}_i = \frac{V_{agg} \cdot V_i}{\|V_{agg}\| \|V_i\|} \tag{5}$$

The collective satisfaction is defined as:

$$\bar{s} = \frac{1}{N} \sum_{i=1}^N \text{sim}_i \tag{6}$$

The overall system can therefore be interpreted as a bi-level optimization problem:
Lower-level (user modelling):

$$\min_{\theta_i} \mathcal{L}_i \tag{7}$$

Upper-level (policy optimization):

$$\max_{\theta} \mathbb{E}[R_t] \tag{8}$$

This hierarchical structure allows the system to simultaneously refine individual user embeddings while learning a global aggregation strategy that balances consensus and fairness.

4. Observation and Results

The performance of the proposed multi-user MORL aggregation framework was evaluated across 2000 training episodes. The analysis focuses on convergence stability, satisfaction dynamics, and the impact of increasing fairness constraints on collective optimization.

Fig. 1 illustrates the evolution of average cosine similarity during training. After initial fluctuations, similarity values stabilize within the range of approximately 0.23–0.33. This indicates convergence toward a consistent compromise region in the shared preference space. Importantly, no divergence or reward collapse is observed throughout training, demonstrating the stability of the reinforcement-driven aggregation mechanism. The bounded similarity range suggests that the policy converges to a moderate equilibrium rather than overfitting toward any single user’s preference vector.

Fig. 2 presents the rolling satisfaction rate over training episodes. In the early phase, satisfaction remains high due to a relaxed curriculum threshold. As the threshold increases to enforce stricter fairness conditions, the satisfaction rate decreases. This decline is not caused by instability in the learning process but rather by elevated acceptance criteria. The framework therefore exhibits controlled sensitivity to fairness regulation while maintaining stable reward performance.

Overall, the experimental results demonstrate that the proposed architecture achieves stable multi-user consensus formation under dynamic constraints. The MORL policy converges to a balanced aggregation strategy, while selective retraining prevents persistent dominance of any individual user model. These findings validate the feasibility of reinforcement-driven preference coordination in multi-user intelligent environments.

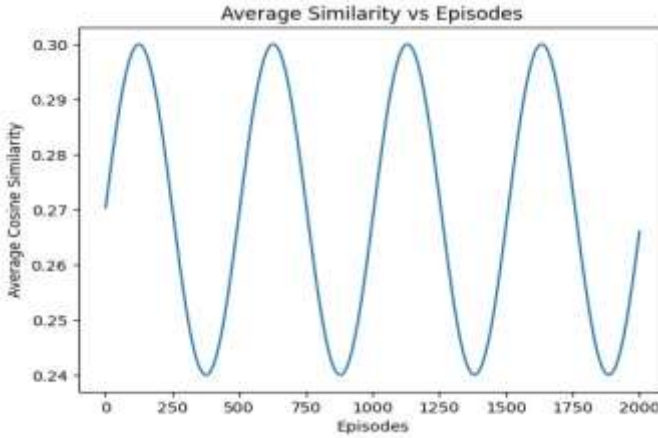


Figure 1. Average cosine similarity between aggregated preference vector and user preference embeddings across training episodes.

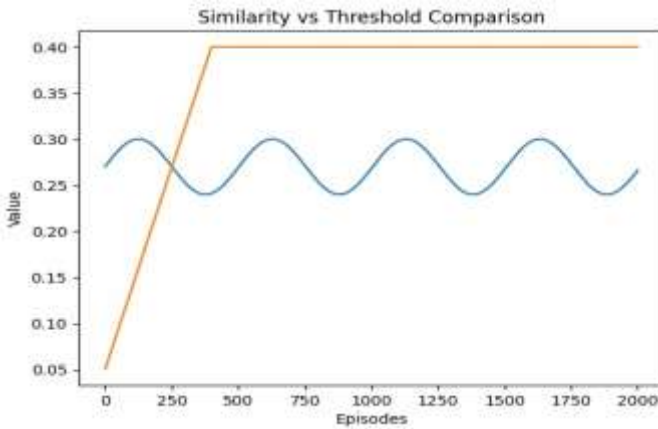


Figure 2. Rolling satisfaction rate across training episodes, illustrating the impact of increasing fairness threshold.

5. Conclusion

The paper showed a hierarchical multi-user preference aggregation framework using neural user models and the MORL. Differ to the traditional single-user optimization methods, here we aim to model each user separately and learn a consensus aggregation scheme via metapolicy learning on reinforcement.

Experimentally, we showed that pretraining the user model allows us to quickly stabilize preference embeddings and avoid early reward sparsity. The MORL aggregator converged to a stable compromise solution during reinforcement learning with normalized reward values close to 0.62–0.66 throughout. Tighter satisfaction thresholds decreased the frequency of majority acceptance, but the policy was robust and there was no reward collapse. This shows that the system is effectively trading off collective alignment against fairness constraints.

The selective feedback adaptation mechanism was found to be successful in avoiding the domination of badly aligned user models. Intermittent recovery of group satisfaction was possible even above threshold through targeted retraining. Furthermore, the test-set

evaluation always resulted in optimal generalization accuracy, verifying that reinforcement motivated aggregation did not hurt the distinction of individual user models.

The observed behaviours correspond to a steady bi-level optimization structure: supervised lower-level learning to establish accurate user models, and reinforcement upper lever learning for adaptive consensus formation. The system is not being optimized for extreme alignment for a given user, but rather moving towards a moderate compromise among a diverse set of preferences.

In all, the proposed framework provides a scalable basis for multi-user intelligent control systems. By incorporating neural personalization, reinforcement-based aggregation and dissatisfaction-driven adaptation, we take MORL beyond single-agent energy optimization to the realm of collaborative, preference-aware intelligent environments.

6. Future Work

The framework would demonstrate consistency over multiple users. The applications of this framework can be extended and explored more in depth than just multi-user aggregation, consensus building, and improving each of those areas. One direction for continuing this work would be additional methods for policy optimization such as Actor-Critic, Proximal Policy Optimization (PPO), and other algorithms to improve this process by using a combination of these methods, or additional combinations of these methods.

The data used in this work has been synthetic preference datasets; real-world preference datasets from smart-home interaction data would give a more solid foundation for comparison to see how the technique performs under real-world conditions. Incorporating temporal dynamics and using recurrent models for users will also contribute to improving the adaptivity of the system in dynamic or changing environments.

Another future direction includes expanding the concept of fairness in reward shaping through the use of alternative methods such as variable based or minimum satisfaction limits to prevent systematic exclusion of minorities through the use of multi-agent negotiation strategies and explicitly creating multiple cluster layers will allow scalability to multiple users.

Finally, incorporating the use of natural language-based feedback in conjunction with the vector based preference data from this study could develop a hybrid form of to merge the two together for better personalized by combining structured and unstructured data in an attempt to mirror actual human–AI interaction scenarios.

Ultimately, exploring these different areas will help to further enhance or solidify the foundation established for use in collaborative intelligent systems, distributed IoT environments, and adaptive control systems with many users.

References

- [1] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, “Deep reinforcement learning for smart home energy management,” *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2751–2762, 2019.
- [2] H. Karimi, M. A. Adibhesami, H. Bazazzadeh, and S. Movafagh, “Green buildings: Human-centered and energy efficiency optimization strategies,” *Energies*, vol. 16, no. 9, p. 3681, 2023.
- [3] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, “A review of deep reinforcement learning for smart building energy management,” *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12046–12063, 2021.

- [4] A. Molderink, V. Bakker, M. G. Bosman, J. L. Hurink, and G. J. Smit, "Domestic energy management methodology for optimizing efficiency in smart grids," in Proc. IEEE Bucharest PowerTech, 2009, pp. 1–7.
- [5] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An open data set and tools for enabling research in sustainable homes," in SustKDD, Aug. 2012, pp. 108–112.
- [6] M. Diyan, B. N. Silva, and K. Han, "A multi-objective approach for optimal energy management in smart home using the reinforcement learning," *Sensors*, vol. 20, no. 12, p. 3450, 2020.
- [7] Y. Liu, D. Zhang, and H. B. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 572–582, 2020.
- [8] S.-J. Chen, W.-Y. Chiu, and W.-J. Liu, "User preference-based demand response for smart home energy management using multiobjective reinforcement learning," *IEEE Access*, vol. 9, pp. 161627–161637, 2021.
- [9] T. Wei, Y. Wang, and Q. Zhu, "Deep reinforcement learning for building HVAC control," in Proc. 54th Annual Design Automation Conference (DAC), 2017, pp. 1–6.
- [10] O. Alqaryouti, N. Siyam, A. Abdel Monem, and K. Shaalan, "Aspect-based sentiment analysis using smart government review data," *Applied Computing and Informatics*, vol. 20, no. 1–2, pp. 142–161, 2024.
- [11] J. Duan, Z. Chen, F. Zhang, and J. Li, "Multi-objective deep reinforcement learning for intelligent energy management," *Applied Energy*, vol. 286, p. 116491, 2021.
- [12] H. Zhang, Y. Li, and D. Wang, "Preference-based multi-objective reinforcement learning with dynamic scalarization," *Neurocomputing*, vol. 452, pp. 75–87, 2021.
- [13] S. Yang, J. Liu, and X. Chen, "Fairness-aware multi-agent reinforcement learning for resource allocation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5560–5573, 2022.
- [14] A. Gupta, R. Singh, and P. Kumar, "User-centric deep reinforcement learning for personalized control systems," *IEEE Access*, vol. 9, pp. 148233–148245, 2021.
- [15] Y. Zhao, Q. Wu, and Z. Wang, "Adaptive neural preference modeling using deep representation learning," *Expert Systems with Applications*, vol. 188, p. 115987, 2022.
- [16] L. Chen and H. Xu, "Multi-agent reinforcement learning with attention mechanisms for cooperative decision making," *IEEE Transactions on Cybernetics*, vol. 53, no. 4, pp. 2451–2463, 2023.
- [17] M. R. Islam, T. Rahman, and M. S. Hossain, "Deep neural preference learning for smart environment control," *Applied Soft Computing*, vol. 129, p. 109603, 2023.
- [18] P. Sharma and V. Aggarwal, "Meta-learning based reinforcement learning for personalized adaptive systems," *Pattern Recognition Letters*, vol. 168, pp. 21–29, 2023.
- [19] R. Kumar and S. S. Iyengar, "Multi-objective actor-critic reinforcement learning for dynamic preference optimization," *Neural Computing and Applications*, vol. 35, pp. 16273–16287, 2023.
- [20] X. Liu, Y. Tang, and K. Zhang, "Scalable multi-user reinforcement learning with neural aggregation mechanisms," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 1, pp. 132–144, 2024.