Department of Industrial and Systems Engineering
UNIVERSITY OF WISCONSIN–MADISON

Optimizing Team Assignments by Preferences

Kasper Veje Jakobsen, kjakobsen@wisd.edu[1]

[1]Department of Industrial and Systems Engineering

Problem Statement

Optimal team formation is crucial in domains like work places and sports. Each individual has preferences for working with others, which influence team performance. The main challenges are:

- **Unknown Preferences:** Individual preferences are not directly observable.
- **Dynamic Evolution:** Preferences evolve continuously towards all individuals.
- **Complexity:** Large action space (all possible team assignments) and continuous state space (preferences and uncertainties).

Objective: Develop a framework to assign individuals to teams dynamically while:

- **Maximizing** overall **team performance**.
- Balancing **exploration** (learning preferences) and **exploitation** (using known preferences).

Mathematical Notation

Individuals: $I = \{1, 2, \ldots, n\}$: Set of n individuals.

Team Assignments: At each period p, individuals are grouped into mutually exclusive teams:

$$\mathcal{T}^p = \{T_1^p, T_2^p, \ldots, T_{s_p}^p\}, \quad T_j^p \cap T_k^p = \emptyset, \quad \cup_{j=1}^{s_p} T_j^p = I.$$

Preferences: The true (unobserved) preference w_{ij}^p evolves continuously:

$$w_{ij}^0 = N(0, \sigma_p^2), \quad w_{ij}^{p+1} = w_{ij}^p + \varepsilon_{ij}^p, \quad \varepsilon_{ij}^p \sim N(0, \sigma_w^2)$$

Feedback: Sparse feedback F_{ij}^p is collected for individuals in the same team:

$$F_{ij}^p = w_{ij}^p + \eta_{ij}^p, \quad \eta_{ij}^p \sim N(0, \sigma_f^2)$$

DP Formulation

State Space: The state at each period represents the belief about preferences, expressed as a probability distribution:

$$\hat{s}_p = \left\{ P(w_{ij}^p \mid \text{history up to } p) \mid i, j \in I \right\}.$$

This belief can be fully characterized by the current estimates of preferences and their uncertainties, where μ_{ij}^p is the estimated preference and σ_{ij}^p is the uncertainty:

$$s_p = \{\mu_{ij}^p, \sigma_{ij}^p \mid i, j \in I\},$$

Action Space: The action $a_p \in \mathcal{A}$ is a valid team assignment for the period.

Transition Function: Beliefs about the preferences evolve based on feedback:

$$s_{p+1} = f(s_p, F^p), \quad \text{where } F^p \text{ incorporates feedback on evolving preferences.}$$

Reward Function: The reward reflects team performance, measured by the beliefs:

$$r(s_p, a_p) = \sum_{T \in a_p} \left(\sum_{i,j \in T, i \neq j} \mu_{ij}^p \right)$$

Objective: Find the policy π^* that maximizes the expected cumulative reward:

$$\pi^* = \arg\max_{\pi} \mathbb{E} \left[\sum_{p=1}^{m} r(s_p, a_p) \right].$$

Learning Preferences: Kalman Filter

Overview: Using a Kalman filter, the preferences between individuals are dynamically estimated by updating beliefs based on feedback received after team interactions, using $\hat{\sigma}_w$ and $\hat{\sigma}_f$ which are estimates of the process noice and feedback noise. An example of the learning process is shown in Figure 1.

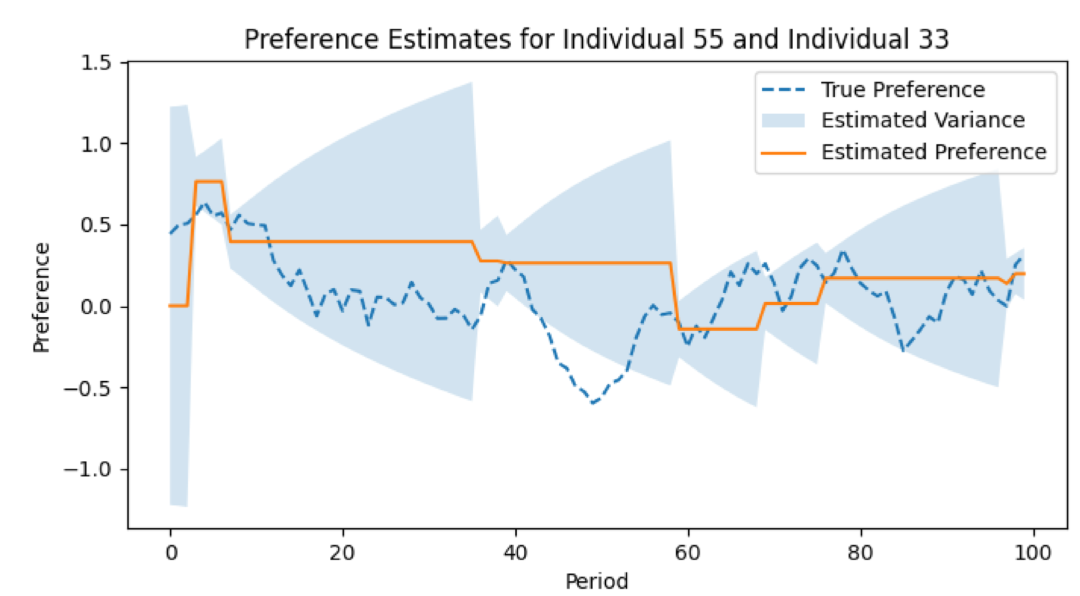

Figure 1. Estimated preferences μ_{ij} compared to true preferences w_{ij}. σ_{ij} reflects uncertainty.

Optimal Assignments

Objective: Assign individuals to teams to maximize overall performance by balancing **exploration** (learning preferences) and **exploitation** (using known preferences).

Upper Confidence Bound (UCB) Strategy: Combines mean preference μ_{ij}^p and uncertainty σ_{ij}^p into a score S_{ij}^p using an exploration weight $\beta \geq 0$:

$$S_{ij}^p = \mu_{ij}^p + \beta \cdot \sigma_{ij}^p.$$

Teams are assigned to maximize the sum of scores:

$$a_p = \arg\max_{a \in \mathcal{A}} \sum_{T \in a} \sum_{i,j \in T, i \neq j} S_{ij}^p.$$

MILP: Used as a baseline, MILP directly optimizes team assignments by maximizing the score (S_{ij}^p) at each period.

Future Directions: Explore scalable methods like multi-agent reinforcement learning (MARL) to efficiently handle large action spaces and dynamic team assignment scenarios.

Integrated Framework

The integrated framework, which uses a *learner* for preference estimation, an *actor* for team assignments, and the *environment* for updates and feedback is shown in Figure 2.

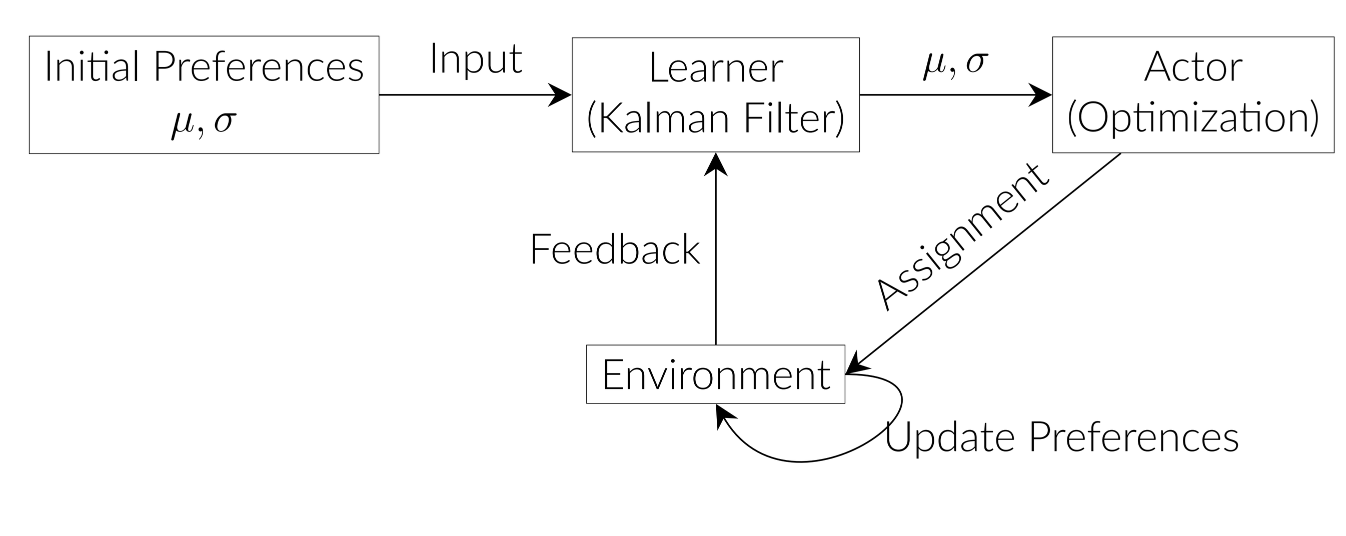

Figure 2. Integration of Learner, Actor, and Feedback Loop with Preference Update in the Environment.

Simulation Setup and Results

Setup: The simulation evaluates team assignment strategies using $n = 10$ individuals, $m = 100$ periods, and $s = 3$ teams. Preferences evolve with noise ($\sigma_w = 0.1$), and feedback is noisy ($\sigma_f = 0.1$). A **Random Assignment** strategy and **Optimal Assignments** with varying exploration weights β are compared. Results are averaged over 10 simulations, with the mean and variance of performances shown in Figure 3 and Figure 4.

Metrics: The reward measures team performance over periods, reflecting the effectiveness of assignments. The preference distance evaluates the convergence of estimated preferences (μ_{ij}) to the true preferences (w_{ij}) using the L_2-norm, indicating the learning accuracy of the model.

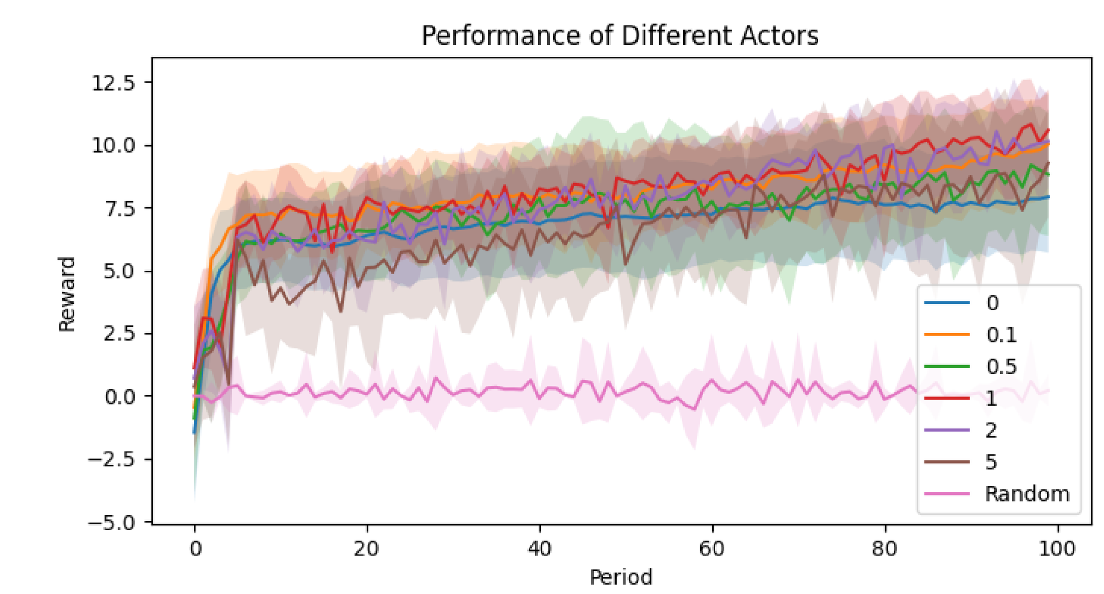

Figure 3. Performance comparison across strategies. Higher rewards indicate better team assignments.

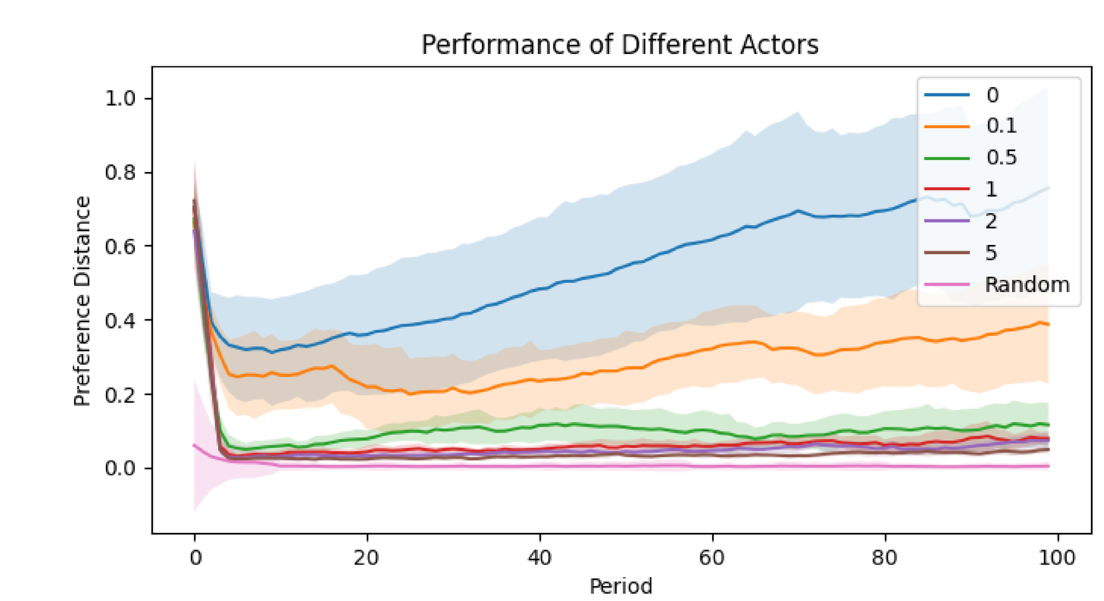

Figure 4. Preference distance between estimated and true preferences. Lower values show improved learning accuracy.

Discussion and Conclusion

The results demonstrate that incorporating both **exploration** (via uncertainty) and **exploitation** (via mean preferences) significantly enhances team performance compared to random assignments. The Kalman filter effectively learns individual preferences over time, with convergence improving as the number of interactions increases.

The optimization-based approach balances learning and performance, but computational complexity limits its scalability to larger settings. Future work will explore **scalable methods** such as reinforcement learning or heuristic-based optimization to handle larger action spaces.

This framework provides a foundation for dynamic team assignment, with potential applications in corporate, academic, and sports environments where collaboration dynamics evolve over time.