Department of Industrial and Systems Engineering

UNIVERSITY OF WISCONSIN–MADISON

Optimizing Team Assignment

Kasper Veje Jakobsen, kjakobsen

[1]Department of Industrial and Systems En

ts by Preferences

@wisd.edu[1]

gineering

Problem Statement

Optimal team formation is crucial in domains like work places and sports. Eacl
preferences for working with others, which influence team performance. The r
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 continuo
(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 p

Mathematical Notation

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

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

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

n individual has
main challenges

ous state space

e:

preferences).

sive teams:

Learning Preferences: Kalman

Overview: Using a Kalman filter, the preferences between individ
by updating beliefs based on feedback received after team intera
are estimates of the process noice and feedback noise. An exa
shown in Figure 1.

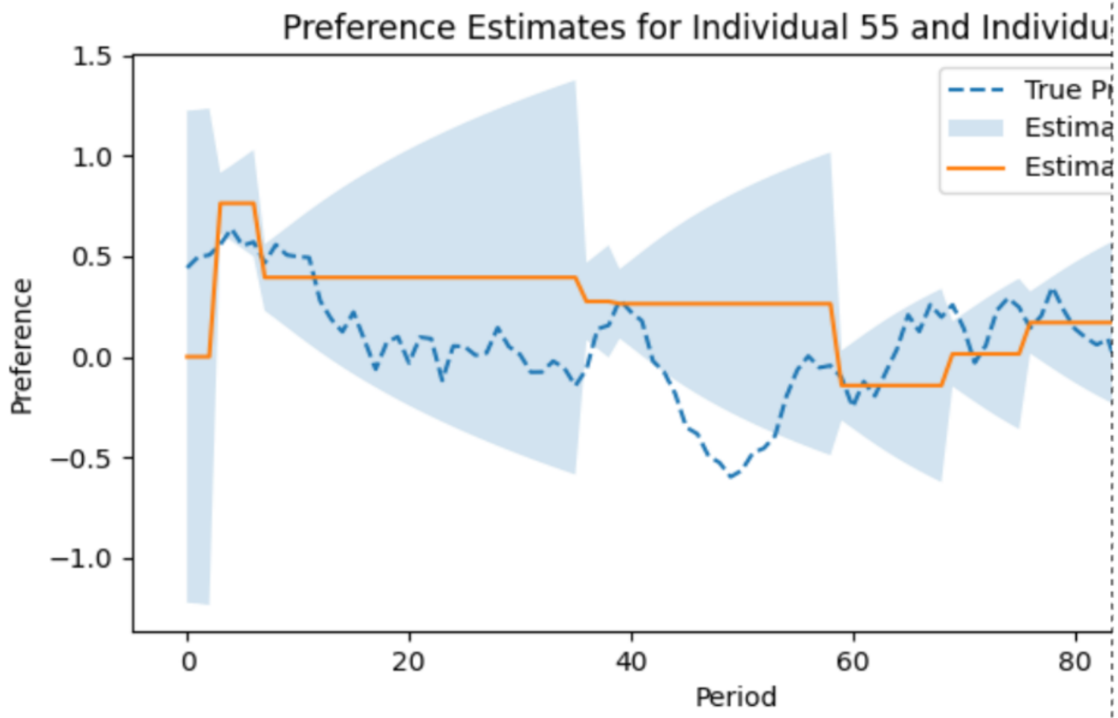
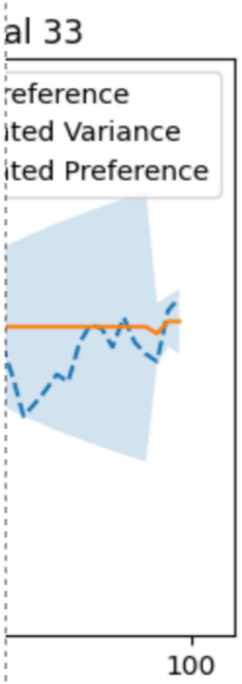


Figure 1. Estimated preferences μ_{ij} compared to true preferences u

Optimal Assignments

Filter

luals are dynamically estimated
actions, using $\hat{\sigma}_w$ and $\hat{\sigma}_f$ which
mple of the learning process is

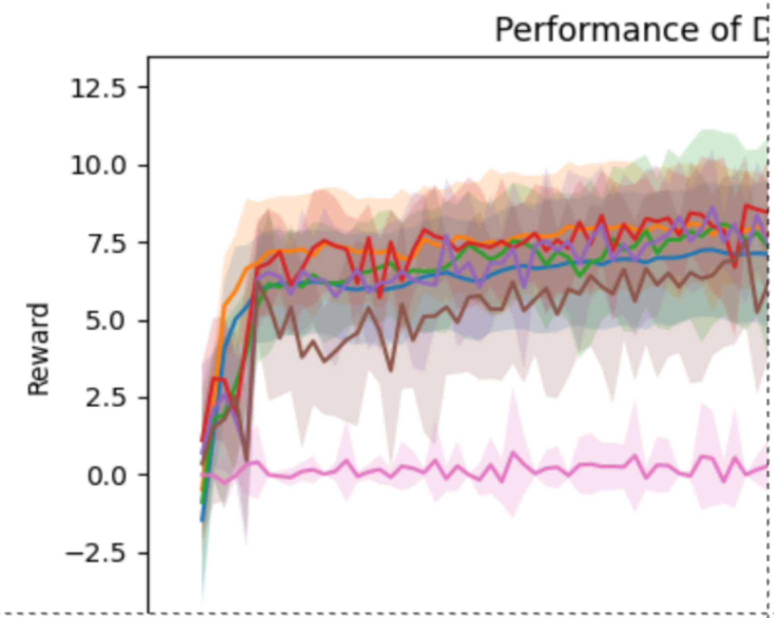


$_{ij}$. σ_{ij} reflects uncertainty.

Simulation Setu

Setup: The simulation evaluates team assignmen
periods, and $s = 3$ teams. Preferences evolve
$(\sigma_f = 0.1)$. A **Random Assignment** strategy and **(**
weights β are compared. Results are averaged ov
of performances shown in Figure 3 and Figure 4.

Metrics: The reward measures team performanc
assignments. The preference distance evaluates t
to the true preferences (w_{ij}) using the L_2-norm, ir

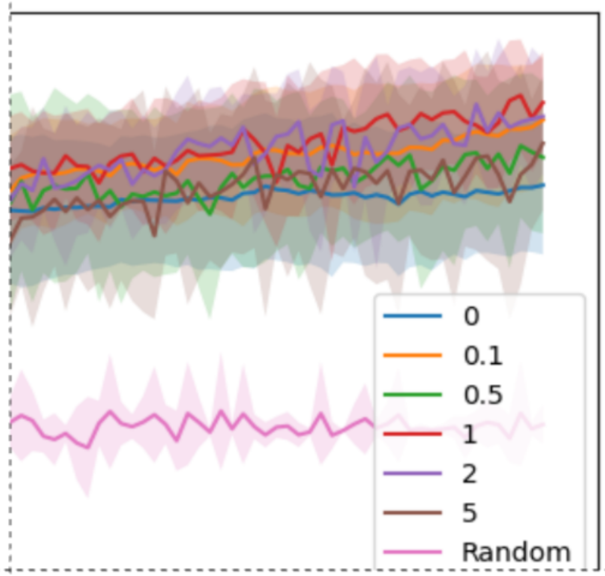


p and Results

t strategies using $n = 10$ individuals, $m = 100$ with noise ($\sigma_w = 0.1$), and feedback is noisy **Optimal Assignments** with varying exploration er 10 simulations, with the mean and variance

e over periods, reflecting the effectiveness of he convergence of estimated preferences (μ_{ij}) ndicating the learning accuracy of the model.

ifferent Actors



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, 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 ties, 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:

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

expressed as a

their uncertain-

ences.

Objective: Assign individuals to teams to maximize overall perfor
(learning preferences) and **exploitation** (using known preferences

Upper Confidence Bound (UCB) Strategy: Combines mean pref
into a score S^p_{ij} using an exploration weight $\beta \geq 0$:

$$S^p_{ij} = \mu^p_{ij} + \beta \cdot \sigma^p_{ij}.$$

Teams are assigned to maximize the sum of scores:

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

MILP: Used as a baseline, MILP directly optimizes team assignmen
at each period.

Future Directions: Explore scalable methods like multi-agent rei
efficiently handle large action spaces and dynamic team assignme

Integrated Framework

The integrated framework, which uses a *learner* for preference es
signments, and the *environment* for updates and feedback is show

nance by balancing **exploration**
).

erence μ_{ij}^{p} and uncertainty σ_{ij}^{p}

ts by maximizing the score (S_{ij}^{p})

nforcement learning (MARL) to
ent scenarios.

stimation, an *actor* for team as-
vn in Figure 2.

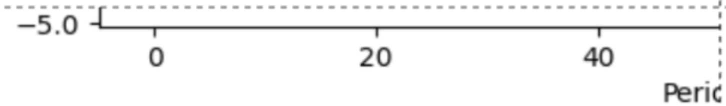


Figure 3. Performance comparison across strategies.

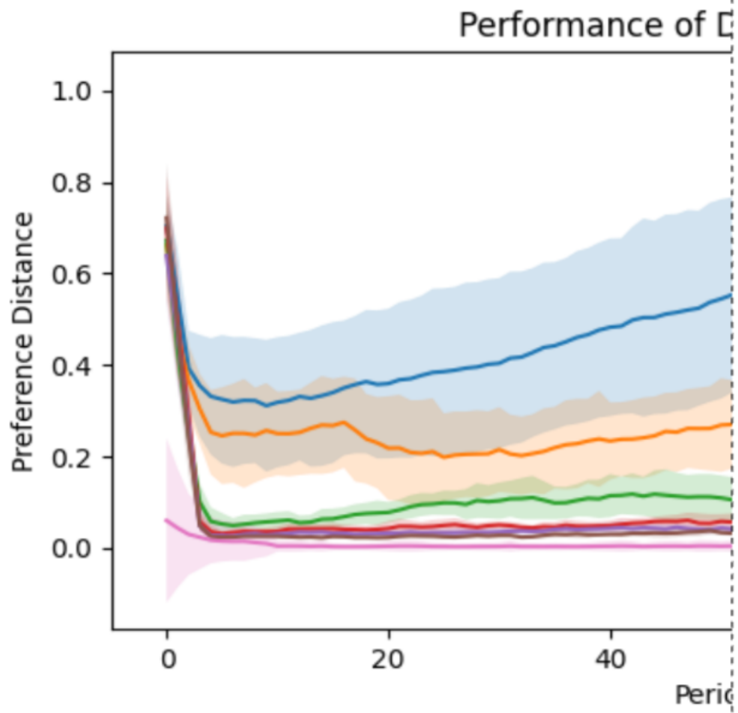
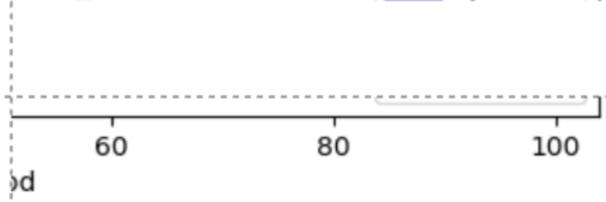
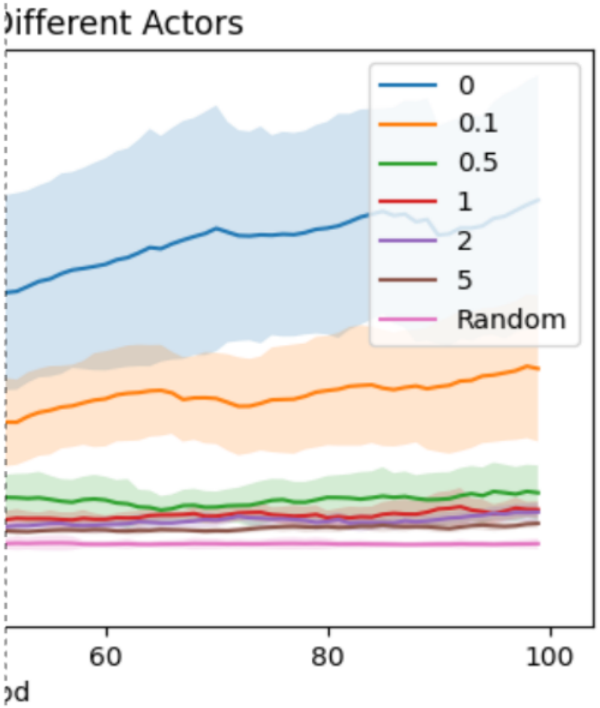


Figure 4. Preference distance between estimated and true p
accuracy.

Discussion and



Higher rewards indicate better team assignments.



›references. Lower values show improved learning

ı Conclusion

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

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

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

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

fs:

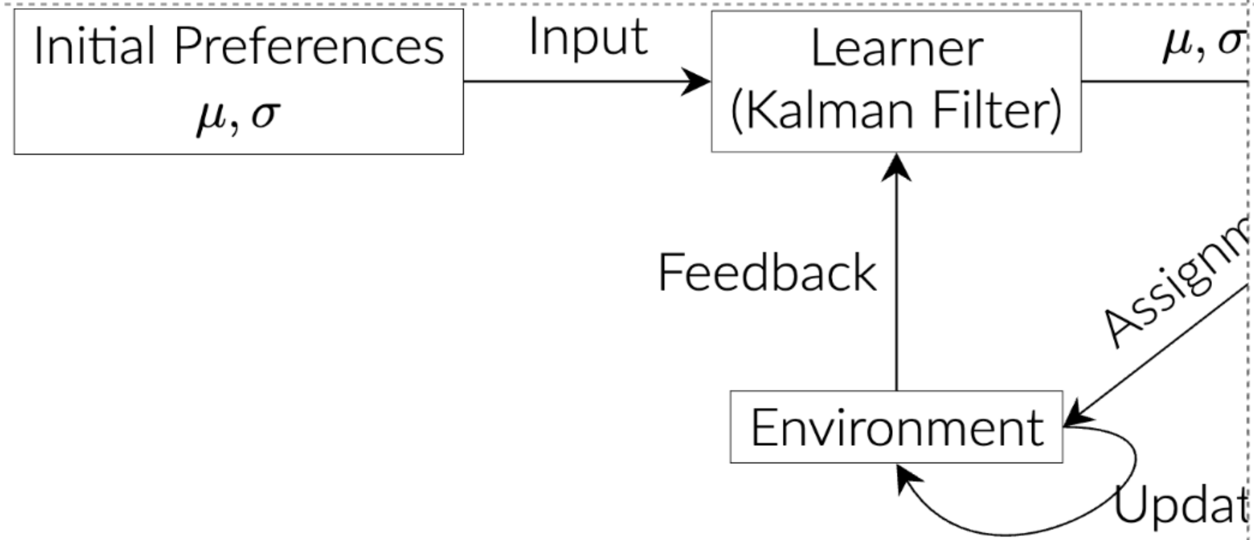
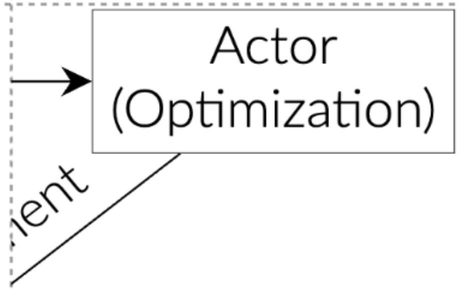


Figure 2. Integration of Learner, Actor, and Feedback Loop with Preferer



te Preferences

nce Update in the Environment.

The results demonstrate that incorporating both **ex**
mean preferences) significantly enhances team pe
The Kalman filter effectively learns individual prefe
as the number of interactions increases.

The optimization-based approach balances learni
plexity limits its scalability to larger settings. Futu
reinforcement learning or heuristic-based optimiz

This framework provides a foundation for dynami
in corporate, academic, and sports environments w

-Madison

xploration (via uncertainty) and **exploitation** (via
erformance compared to random assignments.
erences over time, with convergence improving

ng and performance, but computational com-
re work will explore **scalable methods** such as
ation to handle larger action spaces.

c team assignment, with potential applications
where collaboration dynamics evolve over time.