Multi-Stage Vehicle Assignment with Equilibrium Selection Framework

Areeb Zahid [1] **Sadaan Tahir** [2]

Abstract

The vehicle-target assignment problem poses significant challenges in multi-agent systems due to the dynamic interactions between agents and the need to optimize for multiple objectives, such as compatibility, fairness, and system-wide utility. This paper addresses these challenges by developing a hybrid framework that integrates a multi-stage stochastic game with an equilibrium selection mechanism. Vehicles are dynamically categorized into types (light and heavy), while targets are assigned heterogeneous types (weak, medium, and strong), reflecting real-world constraints.

Building on the Unified Learning Framework (Algorithm 1), the proposed method employs a softmax-based action selection policy to guide vehicles toward optimal target assignments. The framework incorporates Bellman updates and transition probabilities to iteratively improve Q-values and state-value functions, ensuring convergence to optimal policies. Compatibility constraints ensure that vehicles are only assigned to suitable targets, while an availability tracker dynamically updates the game state, terminating when all targets are acquired.

Through simulation, the proposed approach demonstrates its scalability by handling up to large number of finite dynamically assigned targets, while maintaining fairness and cumulative utility maximization. The results showcase the efficacy of the framework in achieving optimal and equitable target assignments under resource heterogeneity and real-time constraints, providing a robust solution for autonomous resource allocation problems.

1. Introduction

In multi-agent systems, optimization and coordination problems are often challenging due to the dynamic interactions among agents with potentially conflicting goals. One such problem arises in the context of vehicle-target assignment, where autonomous vehicles must be allocated to various targets while optimizing for performance metrics such as reward maximization, task compatibility, and system-wide utility (Arslan et al. [2007].) This problem becomes increasingly complex when incorporating constraints like resource heterogeneity and the need for equitable resource allocation.

Conventional approaches to vehicle-target assignment often rely on static optimization methods or centralized planning, which can be computationally demanding and infeasible in dynamic or decentralized environments. Furthermore, the lack of adaptability to real-time changes, such as varying target states or evolving vehicle capabilities, limits their effectiveness in real-world applications.

To address these challenges, we propose a hybrid framework that integrates multi-stage stochastic games with equilibrium selection mechanisms (Zhang et al. [2024].) This approach not only enables decentralized decision-making but also ensures optimal assignment by leveraging a structured learning framework. Specifically, we model the problem as a multi-stage game where vehicles, categorized as light or heavy, dynamically assign themselves to targets classified as weak, medium, or strong, based on compatibility constraints. Using a softmax-based equilibrium selection algorithm, we achieve an optimized assignment strategy while ensuring fairness and robustness across agents.

A key innovation in our approach is the integration of equilibrium selection algorithms that prioritize optimality under system constraints, such as avoiding incompatible assignments (e.g., light vehicles attempting strong targets). Additionally, our framework dynamically tracks global utility, ensuring cumulative performance is maximized over multiple stages. By incorporating heterogeneity in vehicle and target types, we simulate a realistic scenario that mirrors challenges in domains such as autonomous surveillance, disaster response, and military logistics.

This paper presents the following contributions:

1. A hybrid framework combining multi-stage stochastic games with equilibrium selection for dynamic vehicle-target assignments.
2. The incorporation of vehicle-target heterogeneity, which enforces realistic constraints and optimizes compatibility.
3. A dynamic utility tracking mechanism that ensures fair

and efficient resource allocation across agents while maximizing cumulative system-wide utility.

4. Comprehensive evaluation through simulations, demonstrating the efficacy of the proposed approach in achieving optimal assignments under varying scenarios.

2. Methodology

In this section, we detail our methods employed to address problem by leveraging concepts from equilibrium selection in multi-agent systems and incorporating a dynamic, heterogeneity-aware framework for decision-making. The approach builds upon the foundational algorithms for equilibrium selection, adapts them to a newly structured vehicle assignment game, and consolidates these methodologies into a hybrid solution.

2.1. Algorithms for Equilibrium Selection

Our approach incorporates the foundational principles of Algorithm 1 (Unified Learning Framework) and Algorithm 2 (Sample-Based Unified Learning Framework) from the equilibrium selection literature. These algorithms serve as the backbone for achieving optimal decision-making in multi-agent systems.

Algorithm 1: Unified Learning Framework (ULF):
Algorithm 1 adopts an Actor-Critic learning structure where:

- The actor selects actions (vehicle-target assignments) based on a stochastic policy derived from Q-values.
- The critic evaluates the outcomes of these actions by updating the Q-values and state-value functions (V).

The Q Value Update Rule is as follows:

$$Q_{i,h}^{t+1}(s,a) = r_{i,h}(s,a) + \sum_{s'} P_h(s'|s,a) V_{i,h+1}^{t+1}(s')$$

Where $r_{i,h}$ is the reward, P_h is the transition probability, and V is the state value.

The V Value Update Rule is as follows:

$$V_{i,h}^{t+1}(s) = \frac{t}{t+1} V_{i,h}^{t}(s) + \frac{1}{t+1} Q_{i,h}^{t}(s, a_h^t(s))$$

The actor's policy updates are guided by a softmax function, ensuring exploration and exploitation. This algorithm provides the theoretical underpinnings for convergence to an optimal equilibrium in the absence of sample-based constraints.

Algorithm 2: Sample-Based ULF:
Algorithm 2 extends Algorithm 1 by introducing sample-based updates for improved computational efficiency:

- The algorithm estimates Q-values using trajectories sampled from the system's state-space.
- A visitation count $N_h(s,a)$ is maintained to dynamically adjust the learning rate for Q-value updates:

$$Q_{i,h}^{t+1}(s,a) = \frac{N_h(s,a)-1}{N_h(s,a)} Q_{i,h}^{t}(s,a) + \frac{1}{N_h(s,a)} [r_{i,h}(s,a) + V_{i,h+1}^{t+1}(s')]$$

This sample-based approach ensures adaptability to real-time decision-making scenarios where full system knowledge may not be available.

2.2. The Vehicle Assignment Game: New Setup

The traditional vehicle assignment problem involved static, deterministic optimization with uniform vehicles and targets. Our proposed setup introduces a more realistic and dynamic framework by incorporating heterogeneity in both vehicles and targets, as well as compatibility constraints.

Heterogeneous Vehicles and Targets:
Vehicles are classified as:

- Light: Best suited for weak or medium targets.
- Heavy: Best suited for strong or medium targets.

Targets are classified as:

- Weak: High utility for light vehicles, no utility for heavy vehicles.
- Medium: Moderate utility for both light and heavy vehicles.
- Strong: High utility for heavy vehicles, no utility for light vehicles.

Dynamic Decision-Making:
This problem is modeled as a multi-stage stochastic game, where:

- Each stage corresponds to a decision-making round.
- Vehicles dynamically assign themselves to available targets while maximizing their individual and collective utilities.
- A softmax-based policy ensures exploration and adaptability in assignment decisions.

Constraints and Termination:
A target, once assigned, becomes unavailable for subsequent stages. The game ends when all targets are acquired or the maximum number of stages is reached.

2.3. Consolidating Equilibrium Selection with the Vehicle Assignment Game

To solve the assignment problem, we integrate the principles of equilibrium selection with the newly defined vehicle-target assignment setup. This hybrid framework consolidates the learning and optimization processes as follows:

Action Selection with Compatibility Constraints:

- Algorithms 1 and 2 provide the Q-value update mechanisms and softmax-based action selection for vehicles.
- A compatibility function enforces that only compatible vehicle-target assignments are allowed:

$$Utility(v,t) = \begin{cases} Reward(v,t) & \text{if compatible} \\ 0 & \text{otherwise} \end{cases}$$

- Vehicles skip incompatible assignments, ensuring optimality and feasibility.

Dynamic Utility Tracking:
Q-values and utilities are updated dynamically based on the cumulative rewards over stages. Vehicles retain their utility from the previous stage unless they acquire a compatible target.

Softmax-Based Equilibrium Selection:

A softmax-based policy is used to model the actor's decision-making:

$$\pi(a|s) = \frac{e^{Q(s,a)/\tau}}{\sum_{a'} e^{Q(s,a')/\tau}}$$

Where τ is the temperature controlling exploration vs. exploitation.

State Transition Probabilities:
The state transition probabilities are generated by initially creating random uniform transition probabilities for each vehicle and target, resulting in a three-dimensional array. To ensure that these probabilities are valid, they are normalized to ensure that they sum to 1 for each combination of state s and action a. This normalization is crucial because it guarantees that the probabilities accurately represent the likelihood of transitioning from one state to another given a specific action.

Global Optimization:
The framework tracks global utility (cumulative system reward) across stages, ensuring the collective performance of vehicles is maximized. Vehicles are incentivized to select targets that improve system-wide efficiency while adhering to compatibility constraints.

3. Results

In this section, we detail the results of our simulation for the multi-stage assignment game with the following parameters set in place:

1. Number of Targets: 200
2. Number of Vehicles: 25
3. Number of Stages: 20

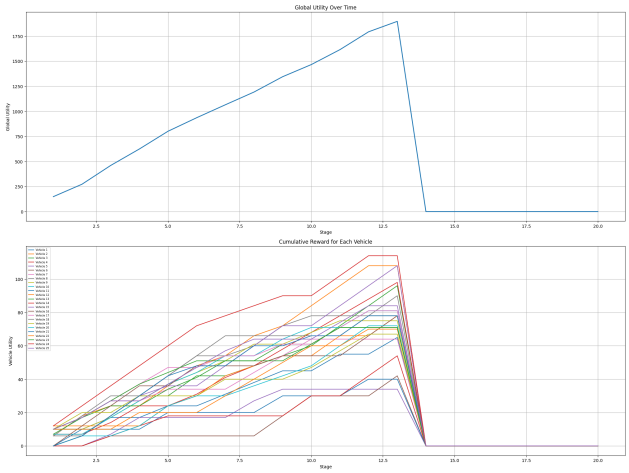

Figure 1. Global & Individual Utilities over each stage.

We chose this arbitrarily, our simulations were run on a larger variation of parameters, yielding similar results.

4. Discussion

The methodology of this multi-agent game is both comprehensive and efficient, emphasizing the dynamic interaction between vehicles and targets over a set time horizon. By initializing parameters such as the number of stages (H), vehicles (N), and targets (S), alongside critical factors like the discount factor (gamma) and the temperature for the softmax policy, the framework sets a robust foundation. Vehicles and targets are randomly assigned types, creating a dynamic reward matrix that reflects their compatibility. Key variables, including Q-values and V-values, are initialized to capture expected and maximum rewards, respectively. Additionally, transition probabilities are generated to simulate state transitions, and a compatibility check ensures that only valid vehicle-target pairs are considered for action selection.

Throughout the game, the main loop iterates over each stage, with vehicles selecting actions based on a softmax policy that considers both Q-values and target availability. When a vehicle selects a valid and compatible action, it receives a

reward, which is accumulated into its utility, and the target is marked as unavailable for future assignments. Q and V values are continuously updated to incorporate new information, leveraging probabilistic transitions and the discount factor. The algorithm's efficiency is highlighted by its ability to update global utility after each stage and monitor Q-value convergence. If the Q-values converge, indicating that the game has reached equilibrium, the process concludes early, showcasing the algorithm's effectiveness.

The results of implementing this algorithm with 25 vehicles, 200 targets, and 20 stages demonstrated impressive efficiency. Optimal target assignments for all vehicles were achieved after just 15 stages, significantly faster than initially expected. This rapid convergence underscores the algorithm's capability to manage complex multi-agent interactions swiftly and accurately. By utilizing the structured approach of Algorithm 1, the game ensures that vehicles maximize their rewards through informed action selection and utility updates, maintaining a balance between exploration and exploitation.

5. Conclusion

In conclusion, the proposed multi-agent system efficiently manages vehicle-target assignments through dynamic decision-making, leveraging Q-values, V-values, and softmax policies. The algorithm demonstrates rapid convergence and optimal performance, effectively balancing exploration and exploitation in complex environments, making it well-suited for large-scale, real-time applications.

6. Contributions

- Areeb: 50%
- Sadaan: 50%

References

[1] Zhang, R., & Li, N. (2024). Equilibrium Selection for Multi-agent Reinforcement Learning: A Unified Framework.

[2] Arslan, G. (2007). Autonomous Vehicle-Target Assignment: A Game-Theoretical Formulation.

[3] Rutger Claes, Tom Holvoet, and Danny Weyns. A decentralized approach for anticipatory vehicle routing using delegate multiagent systems. IEEE Transactions on Intelligent Transportation Systems, 12(2):364–373, 2011.

[4] Dean Foster and Peyton Young. Stochastic evolutionary game dynamics. Theoretical population biology, 38(2):219–232, 1990.

[5] David M. Frankel, Stephen Morris, and Ady Pauzner. Equilibrium selection in global games with strategic complementarities. Journal of Economic Theory, 108(1):1–44, 2003. ISSN 0022-0531. doi: https://doi.org/10.1016/S0022-0531(02)00018-2. URL https://www.sciencedirect.com/science/article/pii/S0022053102000182.

[6] Chi Jin, Qinghua Liu, Yuanhao Wang, and Tiancheng Yu. V-learning–a simple, efficient, decentralized algorithm for multiagent rl. arXiv preprint arXiv:2110.14555, 2021.

[7] Marc Lanctot, Vinicius Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. A unified game-theoretic approach to multiagent reinforcement learning. arXiv preprint arXiv:1711.00832, 2017.

[8] Bary SR Pradelski and H Peyton Young. Learning efficient nash equilibria in distributed systems. Games and Economic behavior, 75(2):882–897, 2012.

[9] Buşoniu, L., Babuška, R., & De Schutter, B. (2010). Multi-agent Reinforcement Learning: An Overview. In Studies in computational intelligence (pp. 183–221). https://doi.org/10.1007/978-3-642-14435-6-7

[10] Murphey, R. A., 1999, "Target-Based Weapon Target Assignment Problems," Nonlinear Assignment Problems: Algorithms and Applications, Pardalos, P. M., and Pitsoulis, L. S., ed., pp. 39–53, Kluwer, Dordrecht.

[11] Fudenberg, D., and Levine, D. K., 1998, The Theory of Learning in Games, MIT Press, Cambridge, MA.

[12] Hofbauer, J., and Sandholm, B., 2002, "On the Global Convergence of Stochastic Fictitious Play," Econometrica, 70, pp. 2265–2294.

[13] Fudenberg, D., and Levine, D., 1998, "Learning in Games," European Economic Review, 42, pp. 631–639

[14] Bertsekas, D., and Gallager, R., 1992, Data Networks, 2nd ed., Prentice-Hall, Englewood Cliffs., NJ.

[15] Hart, S., and Mas-Colell, A., 2000, "A Simple Adaptive Procedure Leading to Correlated Equilibrium," Econometrica, 685, pp. 1127–1150