Benchmarking Ant Colony Optimisation and Proximal Policy Optimisation for the Capacitated Dynamic Vehicle Routing Problem with Stochastic Customer Requests

Your Name
University of Cape Town
Cape Town, South Africa
MRXJAY001@myuct.ac.za

ABSTRACT

The Dynamic Vehicle Routing Problem is an extension on the existing classical vehicle routing problem. However, the Vehicle Routing Problem is static and everything is known prior to the computation and does not apply well with the real world of logistics and classic real-world scenarios. The DVRP solves this problem by adding the dynamic aspect of stochastic customers. This research benchmarks two different approaches to solving CDVRP: Ant Colony Optimisation, a metaheuristic algorithm, and Proximal Policy Optimisation, a deep reinforcement learning algorithm. Using Solomon's instances as benchmarks, this will aim to evaluate both algorithms under varying sizes and parameters. Both of these solutions will be assessed on their solution quality, computational time, feasibility and scalability. This will provide insight into the strengths and weaknesses of both these algorithms to evaluate their practical applicability for solving complex and dynamic real-world problems.

KEYWORDS

Vehicle Routing Problem, Dynamic VRP, Ant Colony Optimisation, Proximal Policy Optimisation, Reinforcement Learning, Metaheuristics

1 INTRODUCTION

- Define the classical VRP and explain its limitations for real-time logistics/dynamics.
- Introduce DVRP and the capacitated and its stochastic variant (CDVRP).
- Describe how stochastic requests and their place in the DVRP.
- Motivation behind the choice to compare ACO and PPO as contrasting approaches.
- Briefly discuss expectations

2 RELATED WORKS

- Summarise the early applications of ACO to VRP and its dynamic variants, including pheromone reinitialisation and immigrant migration strategies/elite solutions.
- Discuss PPO and other RL algorithms for routing problem, discuss stability

3 MATERIALS AND METHODS

3.1 ACO

- Artificial ants build routes using pheromone trails and heuristic information.
- Discuss adjustments for dynamics i.e pheromone evaporation, reinitialisation after change, and immigrant ants.
- Handling stochastic requests: integrate new customers from a Poisson process/lamba function
- Parameter tuning

3.2 PPO

- Policy gradient reinforcement learning with clipped surrogate objective for stable updates.
- CDVRP modelled as a Markov Decision Process:
  - **State:** vehicle positions, unserved requesters or customers, capacity
  - **Action:** next location to visit.
  - **Reward:** minimise travel distance and lateness penalties.
- Network design: shared encoder for route state, separate policy and value heads.
- Simulation environment: dynamic requests added mid-episode via Poisson arrivals.

3.3 Dataset

- Solomon benchmark instances (25, 50, 100 customers).
- Dynamic extension: 70% of customers static, 30% dynamic arrivals.
- Multiple seeds per instance for robustness.

4 RESULTS AND DISCUSSION

- Comparison of ACO and PPO:
  - Total distance.
  - Number of vehicles.
  - dynamic requests served.
  - computational time.
- Feasibility.
- Scalability.
- Visualisation analysis:
  - Tables of metrics.
  - Line plots showing performance vs. problem size.
  - Example routing diagrams for each algorithm.
- Discussion of when each method is most effective.

5 CONCLUSION

- Summarise which algorithm performs best for the instances.
- Relate results to real-world application.
- Suggest future work