Integrated Q-Learning with Firefly Algorithm for Transportation Problems

K.R. Pratiba[1], S. Ridhanya[2], J. Ridhisha[3] and P. Hemashree[4, *]

[1,2,3,4]Department of Artificial Intelligence and Machine Learning, Coimbatore Institute of Technology, Coimbatore, TamilNadu, India

Abstract

The study addresses the optimization of land transportation in the context of vehicle routing, a critical aspect of transportation logistics. The specific objectives are to employ various meta-heuristic optimization techniques, including Genetic Algorithms (GA), Ant Colony Optimization (ACO), Firefly Algorithm (FA), Particle Swarm Optimization (PSO), and Q-Learning reinforcement algorithm, to find the optimal solutions for vehicle routing problems. The primary aim is to enhance the efficiency and effectiveness of land transportation systems by minimizing factors such as travel distance or time while adhering to constraints. The study evaluates the advantages and limitations of each algorithm and introduces a novel-based approach that integrates Q-learning with the FA. The results demonstrate that these meta-heuristic optimization techniques offer promising solutions for complex vehicle routing challenges. The integrated Q-learning with Firefly Algorithm (iQLFA) emerges as the most successful approach among them, showcasing its potential to significantly improve transportation optimization outcomes.

*Corresponding author. Email: shreehema256@gmail.com

1. Introduction

Land transportation plays a pivotal role in our daily lives, facilitating the movement of people and goods from one location to another. However, the challenges presented by growing populations and increasing urbanization have given rise to complex issues within the transportation sector. These challenges encompass traffic congestion, suboptimal routing, insufficient infrastructure, and soaring transportation costs. Effectively addressing these issues necessitates the strategic planning and management of land transportation systems. The optimization of such systems, however, proves challenging due to the dynamic and intricate nature of the problem. Conventional optimization techniques like linear programming and mathematical modeling often fall short in providing optimal solutions for these multifaceted transportation challenges. Thus, there arises a pressing need for innovative and efficient techniques capable of delivering optimal solutions.

Meta-heuristic optimization techniques have emerged as promising solutions to the intricate problems posed by land transportation systems. These techniques draw inspiration from principles found in nature, such as evolutionary processes, swarm intelligence, and biological systems. They offer the potential to provide efficient and effective solutions to the intricate issues surrounding land transportation. This paper conducts a comprehensive comparative study of various meta-heuristic algorithms, including ACO [1], GA [2], PSO [3], FA [4], and iQLFA.

2. Related Work

In a study by [5], the authors proposed a neural combinatorial optimization approach for live vehicle routing based on Deep Reinforcement Learning (DRL). This approach leverages deep neural networks and

heuristics. The proposed deep neural network model employs DRL as its training paradigm. Model parameters are trained using stochastic gradient descent and policy gradient methods. The study's findings indicate that, within the constraints of limited computation time—a critical factor in online routing services—the proposed technique can outperform traditional mathematical programming-based solutions. The work presented in [6] focuses on implementing a context based, improved ACO method for designing tourist routes. The study employs a modified ACO approach grounded in game theory and featuring entropy-weighted learning to enhance the precision of optimal solutions for the Traveling Salesman Problem (TSP) using ACO and MMAS. The problem is segmented into smaller subsets to expedite ACO convergence, pheromone updating is utilized to boost optimization capacity, and information is shared among various subpopulations via a coevolutionary approach to prevent ACO from converging to a local optimum. The resulting route is the optimal choice, characterized by its short length and minimal discomfort through picturesque areas. This approach emphasizes the ability to adapt to new settings through independent learning via trial-and-error. The article in [7] presents a FA heuristic approach to solve the TSP. The FA is implemented with modifications in its parameters to suite for solving TSP. The results of the experimentation provide better results when compared to the implementation of TSP instance with Ant Colony Optimization, GA and Simulated Annealing. A comparison study between two deep learning methods, Q-learning and SARSA, is presented in [8]. The paper introduces the Traveling Salesman Problem with Refueling (TSPwR). While Q-learning eliminates the need for prior information in path planning, it does exhibit drawbacks, including delayed convergence and limited generalizability. This led to the emergence of combined Q-learning with other meta-heuristic-based algorithms. In [9], a Q-learning-based PSO method is proposed for path planning in mobile robots. A comparative study of various meta-heuristic algorithms in TSP is conducted [10]. The results indicate that the basic FA, when implemented with minor parametric adjustments, outperforms ACO, SA, and GA in most scenarios. The optimization model in [11] presents an improved ant colony algorithm model based on a path segmentation strategy. The models demonstrate the capability to efficiently handle various traffic characteristics and yield superior optimization results. The study aims to enhance the algorithm's performance in terms of crossing and modification. The enhanced GA exhibits improved performance, affirming the effectiveness of the proposed enhancement. A hybrid model proposed in [12] uses attention encoder and Long Short-Term Memory (LSTM) network decoder to overcome the coordination failure between the vehicles when state-less attention-based decoder is used. The hybrid model which was experimented on min-max Capacitated Vehicle Routing Problem (mmCVRP) improves the solution quality and computational

efficiency over the baseline methods. Research work in [13] is also a hybrid approach called AC2OptGA. This method uses a combination of three algorithms namely modified ACO, 2-opt edge exchange and GA. The combination of these algorithms exploits the strengths in both global and local searches. The proposed approach evaluated on TSPLIB benchmarks for large instances shows better results than M-GELS, a best-known current approach for solving multiple TSP. In [14], current developments in GA, PSO, and ACO for emergency transportation are examined. The paper introduces a novel hybrid Biogeography-Based Optimization (BBO) technique, which outperforms various cutting-edge algorithms. Other approaches to addressing transportation problems involve machine learning and meta-heuristics. In [15], the TSP is solved using K-means clustering combined with the FA, with experimental results showing superiority over other algorithms from the literature. A similar approach is taken in [16], which utilizes the Whale Optimization Algorithm in conjunction with a K-means clustering model to solve an unclustered TSP. This method achieves an optimal solution with the best iterative cost. [17] proposes an end-to-end learning model for coordinated routing of multiple vehicles, capable of handling both heterogeneous and homogeneous fleet scenarios. When compared to existing learning strategies for routing challenges, the proposed model efficiently manages the coordination of multiple vehicles and produces results comparable to robust optimization heuristic approaches. The work presented in [18] the integration of PSO and Q-learning for swarm mobile robots to find the ideal path in an unfamiliar environment. The results of the study were found to perform better than Q-learning and PSO when considered alone. The article in [19] studies the solving of TSP with the application GA. The parameters in GA are identified and set in advance. Also, the crossover and mutation steps in GA are improved to enhance the performance of the problem solving. The optimization model in [20] uses FA and k-means clustering to identify the minimum tour length among a given set of nodes. The methodology consists of three major steps namely clustering the nodes using k-means clustering, finding the optimal path in each of the cluster using FA, and reconnecting all the clusters and returning the path between them. The experiments of this proposed methodology showed promising results compared to the other existing work.

3. Methodology

3.1. Overview

Reinforcement Learning (RL) is a machine learning paradigm focused on learning through interactions with the environment to maximize cumulative rewards. Key components of RL include the agent, environment, states, actions, rewards, and a policy that maps states to actions.

RL algorithms learn by iteratively estimating value functions or policies and updating them based on observed rewards. The goal of RL is to teach the agent an optimal strategy that maximizes expected cumulative rewards. This is typically achieved through an iterative process of exploration and exploitation. The agent explores the environment by taking different actions, learning from observed rewards and state transitions. It gradually refines its policy based on learned information to exploit actions that yield higher rewards.

DRL extends RL by combining RL with deep neural networks to handle high-dimensional state spaces and complex decision-making problems. DRL can also be applied in the context of metaheuristic optimization algorithms. Metaheuristics are iterative optimization methods used to solve complex optimization problems where traditional exact methods may be infeasible or inefficient. DRL can enhance metaheuristics by leveraging its ability to learn from experience and make adaptive decisions. DRL can be combined with traditional metaheuristic techniques to create hybrid algorithms. This involves using the metaheuristic as a search framework and incorporating DRL components to enhance specific aspects of the algorithm. In other words, DRL can be used to learn effective local search operators or guide the selection of search strategies within the metaheuristic. The integration of DRL with metaheuristic optimization algorithms aims to improve the effectiveness, efficiency, and adaptability of metaheuristics by incorporating learning and adaptive decision-making capabilities. By harnessing the power of DRL, these hybrid approaches have the potential to tackle complex optimization problems more efficiently and discover higher-quality solutions.

3.2 Dataset Description

The transportation dataset employed in this study was synthetically generated for experimental purposes. It comprises two fundamental components: a list enumerating the indices corresponding to the sources and destinations, alongside a Numpy array representing the distances between each conceivable source-destination pair. The shape of the distance matrix adheres to the dimensions of the transportation problem at hand, designated as (num_sources, num_destinations). Here, "num_sources" denotes the count of sources, while "num_destinations" signifies the number of destinations. Each individual element within the matrix encapsulates the distance, whether it pertains to actual physical distance, cost, or time, linking a specific source to its associated destination.

It's noteworthy to mention that in practical scenarios, the data sources for transportation problems typically include real-world data such as geographical coordinates, road networks, or historical transportation records. These data sources undergo preprocessing steps to ensure accuracy and suitability for optimization algorithms.

Preprocessing might involve tasks like data cleaning, transforming geographical coordinates into distance metrics, and ensuring data consistency. However, for the purposes of this study, a simplified approach of random data generation was adopted to create a representative dataset for experimentation.

The synthetic distance matrix, encompassing cost or distance values for all source-destination pairs, serves as the foundational input for solving transportation problems. The primary aim is to ascertain the optimal allocation of goods from sources to destinations, all the while accounting for various constraints, including capacity limitations and the imperative to minimize expenses. This distance matrix constitutes an essential component for optimization algorithms, furnishing them with precise cost or distance metrics that guide the iterative search for solutions converging towards the pinnacle of efficiency and cost-effectiveness.

3.3 Algorithms

Ant Colony Optimization. The transportation problem can be resolved using the metaheuristic optimization approach termed ACO. It is a nature-inspired algorithm that emulates the foraging behavior of ants. In the context of solving the transportation problem, ACO consists of several phases. In the construction phase, artificial ants build solutions by selecting routes based on pheromone levels and heuristics, closely mimicking the behavior of real ants in finding paths. After each construction phase, pheromone levels are updated to reinforce the attractiveness of good routes and diminish poor ones. This pheromone update process guides subsequent ant exploration. The algorithm iterates through multiple cycles, and the best solution found during these iterations is retained as the output, representing an optimized transportation plan.

Genetic Algorithm. The transportation problem can be handled through the metaheuristic optimization methodology referred to as the GA. In the context of GA applied to the transportation problem, the methodology begins with the initialization of an initial population of potential vehicle routes, which are randomly generated representations of solutions. Subsequently, the algorithm employs a crossover mechanism, where routes from the population are combined, strategically selecting crossover points to ensure the newly generated routes remain feasible solutions for the transportation problem. To introduce diversity and explore different solution spaces, a mutation operation is applied, causing some routes to undergo alterations in a manner that preserves feasibility. Finally, routes are selected for the next generation based on their fitness, determined by factors such as the total travel distance and their adherence to the problem's constraints. This iterative process continues, with new generations of routes evolving over time, ultimately

converging toward optimal or near-optimal solutions for the transportation problem.

Firefly Algorithm. Transport problems can be overcome using the Firefly methodology, a metaheuristic optimization technique. The FA draws inspiration from the flashing behavior of fireflies in nature and is adapted to solve optimization problems such as the transportation problem. The algorithm starts by initializing fireflies randomly on the solution space, where each firefly represents a potential vehicle route. Fireflies move toward brighter fireflies, and the brightness of a firefly is determined by an objective function, typically aimed at minimizing travel distances. Some fireflies explore by moving randomly to escape local optima. Multiple iterations are performed to allow the algorithm to converge, and the position of the brightest firefly found across these iterations represents the best solution to the transportation problem.

Particle Swarm Optimization. The transportation problem is one of many optimization challenges to which PSO has been applied. It is a population-based optimization technique. In the context of addressing the transportation problem, the algorithm begins with the initialization of particles, each representing a potential vehicle route. Particles adjust their velocities based on their historical best position and the best position within their neighborhood, simulating social interactions and individual learning. These velocity adjustments guide particles in updating their positions within the solution space. Multiple iterations are performed to allow particles to explore the search space, and the best solution found throughout these iterations is retained as the final result.

Q learning. Q-Learning is a reinforcement learning algorithm employed to solve complex decision-making problems, including the transportation problem. The algorithm operates based on a framework involving states and actions. States correspond to the current state of the transportation problem (e.g., vehicle locations), and actions represent potential moves (e.g., selecting a customer to serve). Rewards are assigned to actions based on their quality concerning the current state. Q-values, representing the expected cumulative rewards for taking specific actions from particular states, are updated iteratively using the Q-learning algorithm. As the algorithm progresses, a policy is learned from the Q-values, mapping states to actions, ultimately guiding the agent (e.g., a vehicle) to make optimal decisions within the transportation problem to maximize cumulative rewards.

Integrated Q learning With Firefly Algorithm. The integrated approach of combining Q-learning with the FA for solving the transportation problem involves representing the problem as a Markov Decision Process to facilitate this integration, where states represent relevant information about the system, actions correspond to routing decisions, and rewards quantify the quality of decisions. During each iteration of the FA, fitness evaluation and attractiveness calculations guide fireflies (representing vehicles) toward attractive routes. Simultaneously, Q-learning operates within states to select actions and update a Q-table based on observed rewards and state transitions (using Bellman equation). Through iterative refinement, the integrated algorithm develops a policy that maps states to actions, ultimately determining the best feasible policy to optimize the transportation problem, leveraging the exploration capabilities of the FA and adaptive decision-making of Q-learning for efficient and high-quality solutions.

4. Results and Discussion

The performance evaluation of these optimization algorithms in the context of solving transportation problems was conducted using fitness function values as a crucial metric for assessing the quality of solutions. The fitness function employed was tailored to the specific objectives and constraints of the transportation problem under consideration. The choice of fitness function is paramount, as it directly influences the algorithms' ability to optimize the problem effectively.

In the comparative study of optimization algorithms for solving transportation problems, ACO demonstrated its competitiveness in achieving high-quality solutions, especially in scenarios where routing efficiency and constraint adherence were critical. PSO exhibited robust performance, balancing multiple objectives like minimizing travel distance and vehicle utilization due to its swarm intelligence-based approach. Q-learning, a reinforcement learning technique, showed promise by learning optimal policies over time despite potentially requiring more iterations. GA reliably delivered quality solutions, particularly for large-scale transportation problems, thanks to its population-based exploration. FA excelled in exploratory scenarios with its unique approach inspired by fireflies' behavior.

Notably, the novel approach of iQLFA exhibited outstanding performance compared to traditional methods. However, the standout performer was the iQLFA, which seamlessly merged the FA's exploration capabilities with Q-learning's adaptive decision-making. This integration not only generated high-quality and feasible solutions but also did so with significantly reduced computational time, potentially revolutionizing the field of transportation optimization.

Table 1 illustrates the overall distance of the meta-heuristic methods utilized in this study. The sign indicates that the optimization problem is a minimization problem. The magnitude indicates the distance measure. Q learning algorithm costs 32.90851368, FA costs 39.43798039, PSO costs 76.3181706, GA costs 29.51109282, ACO costs 40.76108214 while the novel method that is proposed in the study, iQLFA costs 22.18895968 which is the least of all.

Figure 1. The comparison plot infers that the iQLFA has obtained less cost when compared with other meta-heuristic algorithms.

Table 1. Total distance values of the various optimization algorithms used in this study.

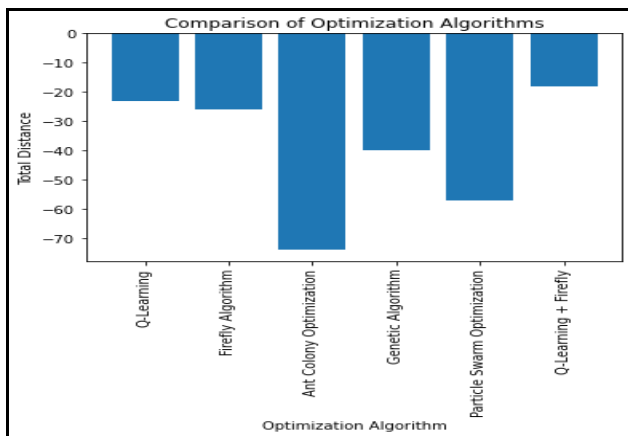| Optimization Algorithms	Total Distance
Q learning	-32.90851368
FA	-39.43798039
PSO	-76.3681706
GA	-29.51109282
ACO	-40.76108214
iQLFA	-22.18895968

Figure 1. Comparison plot for optimization algorithm

5. Conclusion and Future Work

This research has presented an innovative approach, iQLFA, to address transportation optimization problems. A comprehensive comparative study was conducted, evaluating the strengths and limitations of various optimization algorithms, including ACO, PSO, Q-learning, GA, FA, and the proposed iQLFA. The assessment was based on fitness function values that accurately reflect the optimization goals and constraints of transportation problems.

The findings of this study clearly demonstrate the superior performance of the iQLFA over other meta-heuristic optimization techniques. Specifically, the novel approach resulted in the lowest cost among all algorithms considered. These results underline the potential of this integrated approach for addressing complex transportation optimization problems.

Overall, this research contributes to advancing the field of transportation optimization, offering more efficient and effective solutions to transportation planning and management challenges while reducing associated costs and complexities.

5.1 Future Work

This study opens up several avenues for further research and improvement. Future research can explore the application of the proposed approach to different real-time datasets, including various modes of transportation such as air, sea, space, and cable networks. Additionally, metaheuristic algorithms can be employed for hyperparameter tuning of machine learning and deep learning models. This optimization process can fine-tune model parameters, potentially enhancing overall performance and efficiency. It presents an opportunity to comprehensively evaluate system performance and assess alternative models for improved fitness value and efficiency.

Furthermore, investigating the scalability of iQLFA for larger transportation networks is essential. Scalability considerations are crucial for assessing the applicability of the approach to real-world scenarios with extensive transportation systems. Additionally, examining the parallelization of the algorithm can further enhance its efficiency and suitability for larger-scale problems.

In conclusion, this research provides a solid foundation for future advancements in transportation optimization. Subsequent studies can build upon the success of the integrated approach to address even more complex and extensive transportation challenges, contributing to the ongoing improvement of transportation planning and management systems.

References

[1] Marco Dorigo, Christian Blum. Ant Colony Optimization Theory: A Survey. Theoretical Computer Science. 2005; 344(2-3):243-278

[2] David E Goldberg and William Shakespeare. Genetic Algorithms. 2002.

[3] Vijay Kumar, Dinesh Kumar. A Systematic Review on Firefly Algorithm: Past, Present, and Future. Archives of Computing Methods in Engineering. 2020; 28:3269–3291.

[4] James J. Q. Yu, Wen Yu, Jiatao Gu. Online Vehicle Routing with Neural Combinatorial Optimization and Deep Reinforcement Learning. IEEE Transactions on Intelligent Transportation Systems. 2019; 20(10):3806-3817.

[5] Mohamed Ben Ahmed; Farah Zeghal Mansour; Mohamed Haouari. A PSO Approach for Robust Aircraft Routing. In: Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); 06-09 Dec; Singapore: IEEE; 2015 p. 219–223.

[6] A. L. C. Ottoni, E. Nepomuceno, Marcos Santos de Oliveira, Daniela Carine Ramires de Oliveira. Reinforcement learning for the Traveling Salesman

Problem with Refueling. Complex & Intelligent Systems. 2022; 8:2001–2015.

[7] Sharad Kumbharana. Solving Travelling Salesman Problem Using Firefly Algorithm. International Journal for Research in Science & Advanced Technologies. 2013; 2(2):53-57.

[8] Feng Wu. Contactless Distribution Path Optimization Based on Improved Ant Colony Algorithm. Mathematical Problems in Engineering. 2021; 2021:1-11.

[9] Robin T. Bye, Magnus Gribbestad, Ramesh Chandra, Ottar L. Osen. A Comparison of GA Crossover and Mutation Methods for the Traveling Salesman Problem. In: Janusz Kacprzyk. Advances in Intelligent Systems and Computing. Proceedings of Innovations in Computational Intelligence and Computer Vision (ICICV 2020); 17-19 Jan; Manipal University, Jaipur. Springer Singapore; 2020. p. 529-542.

[10] Min-Xia Zhang, Bei Zhang, Yu-Jun Zheng. Bio-Inspired Meta-Heuristics for Emergency Transportation Problems. Algorithms. 2014; 7(1):15-31.

[11] Anahita Sabagh Nejad, Gabor Fazekas. Solving a Traveling Salesman Problem using Meta-Heuristics. International Journal of Artificial Intelligence. 2022; 11(1):41-49.

[12] Aigerim Bogyrbayeva, Taehyun Yoon, Hanbum Ko, Sungbin Lim, Hyokun Yun, Changhyun Kwon. A Deep Reinforcement Learning Approach for Solving the Traveling Salesman Problem with Drone. Transportation Research Part C: Emerging Technologies. 2023; 148:103981.

[13] Youssef Harrath, Abdul Fattah Salman, Abdulla Alqaddoumi, Hesham Hasan, Ahmed Radhi. A Novel Hybrid Approach for Solving the Multiple Traveling Salesmen Problem. Arab Journal of Basic and Applied Sciences. 2019; 26(1):103-112.

[14] Anubhav Kumar Prasad, Dharm Raj Singh, Pankaj. A Genetic Method Using Hybrid Crossover for Solving Travelling Salesman Problem. International Journal of Recent Technology and Engineering. 2019; 8(2):5066-5072.

[15] Yu Huang, Xifan Yao, Junjie Jiang. An Improved Firefly Algorithm for Generalized Traveling Salesman Problem. In: Gerhard Goos, Juris Hartmanis. Lecture Notes in Computer Science. Proceedings of Intelligent Computing Theories and Application. August 12–15, 2021, Shenzhen, China. Springer Cham; 2021. p. 739-753.

[16] James Kennedy and Russell Eberhart. Particle Swarm Optimization. In: ICNN'95 - Proceedings of International Conference on Neural Networks. 27 November 1995 - 01 December; Perth, WA, Australia. IEEE; 1995. p. 1942-1948.

[17] Penggang Gao; Zihan Liu; Zongkai Wu; Donglin Wang. A Global Path Planning Algorithm for Robots using Reinforcement Learning. In: Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO); 6-8 Dec; Dali, China. IEEE; 2019. p. 1693–1698.

[18] Syed Irfan Ali Meerza; Moinul Islam; Md. Mohiuddin Uzzal. Q-Learning Based Particle Swarm Optimization Algorithm for Optimal Path Planning of Swarm of Mobile Robots. In: Proceedings of 1st International Conference on Advances in Science, Engineering and Robotics Technology; 03-05 May; Dhaka, Bangladesh. IEEE; 2019. p. 1–5.

[19] Chutian Sun. A Study of Solving Traveling Salesman Problem with Genetic Algorithm. In: 9th International Conference on Industrial Technology and Management (ICITM); 11-13 Feb; Oxford, UK. IEEE; 2020. p. 307-311.

[20] Ameera Jaradat, Bara'ah Matalkeh, Waed Diabat. Solving Traveling Salesman Problem using Firefly algorithm and K-means Clustering. In: Proceedings of IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT); 09-11 Apr; Amman, Jordan. IEEE; 2019. p. 586-589.