A Comparative Study of Classical Heuristics and Reinforcement Learning for the Traveling Salesman Problem

Liu Zhonglin, the University of Hong Kong

MATH3999: Directed Studies in Mathematics

Supervised by: Prof. Zang Wenan

Abstract

The Traveling Salesman Problem (TSP) is a canonical NP-hard problem in combinatorial optimization. This report details a comparative study between a classical heuristic—the Christofides algorithm—and several Reinforcement Learning (RL) models based on a Pointer Network architecture. We developed and evaluated four distinct methods on randomly generated 30-city TSP instances: (1) the classical Christofides algorithm, (2) a pure RL agent trained from scratch, (3) a hybrid RL agent trained with a REINFORCE policy gradient on Christofides-ordered inputs, and (4) an advanced hybrid RL agent trained with an Actor-Critic algorithm. A robust evaluation over 1000 trials demonstrated the definitive superiority of the Christofides algorithm in both solution quality and speed. The study also revealed that while hybrid RL models significantly outperformed the pure RL model, the choice and stability of the training algorithm are critical, yielding nuanced performance differences.

1 Introduction

The Traveling Salesman Problem (TSP) asks for the shortest possible route that visits a given set of cities exactly once and returns to the origin. Its computational difficulty has made it a benchmark for optimization research, with applications in logistics, circuit design, and DNA sequencing. Traditional approaches involve exact solvers, which are intractable for large instances, and approximation algorithms that provide guaranteed bounds on solution quality.

In recent years, deep reinforcement learning has emerged as a promising, flexible alternative for solving such problems. Unlike classical algorithms, which are often highly specialized, RL agents can learn effective policies from experience with minimal prior knowledge. This project explores this intersection by implementing and rigorously comparing the performance of the classical Christofides algorithm against modern RL agents based on the Pointer Network architecture. The goal is to quantify the trade-offs between the two paradigms and investigate whether hybridizing them can lead to improved performance.

2 Methodology

Four distinct solution methods were developed and evaluated on 30-city TSP instances.

2.1 Classical Heuristic: Christofides Algorithm

A scalable version of the Christofides algorithm was implemented as the classical baseline. To handle 30-city problems efficiently, the minimum-weight perfect matching step, which is computationally expensive, was implemented using a fast greedy algorithm. The final tour was generated via greedy shortcutting of the resulting Eulerian circuit.

2.2 Reinforcement Learning Models

The core of the learning-based approach is an Actor-Critic architecture. The **Actor** is a Pointer Network with an LSTM-based encoder-decoder structure and an attention mechanism, which proposes a tour. The **Critic** is a simple Multi-Layer Perceptron (MLP) that evaluates the state provided by the Actor to stabilize training. Three different RL agents were trained using this architecture.

1. **Pure RL (REINFORCE):** The Actor model was trained from scratch using the REINFORCE policy gradient algorithm. The input was a randomly ordered sequence of city coordinates, and the reward was the negative tour length.
2. **Hybrid RL (REINFORCE):** The Actor was trained using the same REINFORCE algorithm. However, its input was a sequence of city coordinates pre-ordered by a full Christofides tour, providing a high-quality structural "hint".
3. **Hybrid RL (Actor-Critic):** This advanced agent used the same Christofides-ordered input as the REINFORCE hybrid but was trained with the more stable Actor-Critic algorithm. The reward signal was the advantage calculated as (Actual Tour Length - Critic's Value Estimate).

3 Experiments and Results

The final evaluation was conducted by testing each of the four methods on 1000 unique, randomly generated 30-city TSP problems where coordinates were sampled from a unit square. The average tour length and average computation time (inference) were recorded.

The aggregated results are presented in Table 1.

Table 1: Final Robust Average Results (1000 Trials, 30 Cities)

Metric	Christofides	Pure RL	Hybrid (REINFORCE)	Hybrid (Actor-Critic)
Avg. Tour Length	5.3280	11.9893	10.4065	10.3216
Avg. Comp. Time (s)	0.0005	0.0232	0.0228	0.0229

4 Analysis and Discussion

The results from the robust evaluation provide several clear insights.

Solution Quality: The classical Christofides algorithm was decisively superior, finding solutions that were, on average, nearly twice as good as the best RL model. Among the learning-based methods, a clear hierarchy emerged. Both hybrid models significantly outperformed the pure RL model, confirming that seeding the agent with a high-quality input sequence from a classical algorithm is a highly effective strategy. The pure RL agent's performance was the poorest, with an average tour length of 11.9893. Interestingly, the advanced Actor-Critic hybrid (10.3216) only slightly outperformed the simpler REINFORCE hybrid (10.4065), suggesting that the primary performance gain came from the hybrid data strategy rather than the training algorithm itself in this context.

Computational Speed: The Christofides algorithm was faster than the RL models by a factor of approximately 45x. The inference times for all three RL models were comparable, as they share the same underlying Pointer Network architecture. This highlights the significant computational overhead required for a forward pass through a deep neural network compared to a lean, specialized heuristic.

5 Conclusion

This study successfully implemented and compared classical and modern reinforcement learning techniques for the Traveling Salesman Problem. The conclusive finding is that for the standard 30-city TSP, the special-

ized Christofides algorithm remains the superior methodology in both solution quality and computational efficiency.

However, the project also yielded critical insights into the application of RL. The marked success of the hybrid models over the pure RL agent demonstrates the immense value of integrating classical domain knowledge into learning-based systems. The final results suggest that the most promising avenue for future research is not necessarily to replace classical algorithms, but to develop more sophisticated architectures and training methods that can robustly learn to make meaningful improvements upon the strong foundations that classical heuristics provide.