Study on Reinforcement Learning and Its Application for Delivery problem

Pham Hoang Duc Nha; Phan Quoc Minh; Hoang Tuan Anh
FPT University

Abstract

To determine the shortest path are one of the most addressed problems in the field of computer science. Many algorithms are currently available to solve the problems. However, consider between performance and accuracy of algorithms is also the dilemma which need to be solved.
Since exact algorithms are too resources-consuming, improving heuristic solutions to obtain as closest as possible the optimal result yet keeping performance to be within the acceptable threshold is currently the better approach to solve the problems. Our thesis mainly focuses on how using **Reinforcement Learning** and to be more specific, **Ant-Q** would improve the accuracy of heuristic approach. Moreover, we also compare results achieved by Ant-Q with other exact and heuristic algorithms. To address further, we also apply clustering the big solution into smaller chunks as to divide and conquer the big problem performance-wise as well as to fit the current situation of the delivery problem when multiple deliverymen involve in one delivery session.

Introduction

The delivery problem was first formulated in 1930, it is a NP-hard problem which makes an algorithm take more steps than usual to solve. Because this problem was evident long before then, there are countless solutions to this problem up to the present time.
By applying Computer Science, especially Machine Learning, we are gradually introducing a solution that does not require human intervention. The three major aspects of Machine Learning are Supervised Learning, Unsupervised Learning and Reinforcement Learning. The question is why Supervised and Unsupervised Learning can not solve the delivery problem, because of the unstable environmental characteristics of the delivery problem that make them unsuitable. So why choose the Reinforcement Learning, the **Reinforcement Learning** have a mission that is training an agent to solve a task without acquiring any knowledge beforehand, the agent uses trial-errors method to learn so it can adapt to the unstable environment.

Plan Implementation

In order to be able to apply machine learning, especially reinforcement learning to delivery problems, we have to go through the following steps:

- **Step 1**: Study on basic concepts of machine learning
- **Step 2**: Study on Q-learning and Ant-Q
- **Step 3**: Implement the Ant-Q algorithm
- **Step 4**: Prepare standard dataset (TSP and ATSP)
- **Step 5**: Experiment and analyze the results
- **Step 6**: Make a conclusion about the results of the analysis
- **Step 7**: Implement K-Means++ algorithm
- **Step 8**: Combine K-Means++ and Ant-Q

Algorithm

To solve delivery problem, we use Ant-Q algorithm. Based on underlying theory of the **Ant Colony Optimization (ACO)** introduced by Dorigo, Maniezzo and Colorni (Dorigo, 1992; Dorigo, Maniezzo and Colorni 1996; Colorni, Maniezzo and Dorigo, 1991; 1992) and the concept of **Q-learning**, **Ant-Q** is a reinforcement learning method where cooperating agents try to find shortest Hamiltonian tours in a weighted complete graph. By combining the 2 approaches, **Ant-Q** not only improve the accuracy but also prevent itself to converge toward local optimal.

To divide the big delivery problem into smaller delivery problem, we use **K-Means++** algorithm. The **K-Means** algorithm is one of the oldest and most popular clustering algorithm. It is proposed by Lloyd and still very widely used today. However, the result of algorithm can be arbitrarily bad compared to the optimal clustering. The **K-Means++** algorithm addresses that problem by choosing random starting centers with very specific probabilities before proceeding with the standard k-means optimization iterations.

Data Analysis

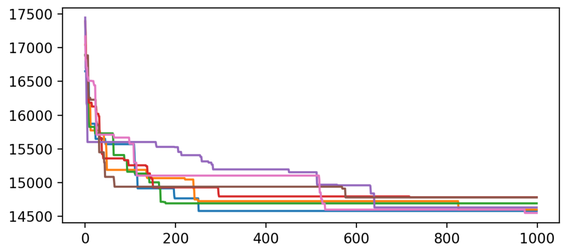

Chart 1. Ant-Q results in Problem:ry48p, 7 trials, 1000 iteration each. All results are in [14500, 15000]

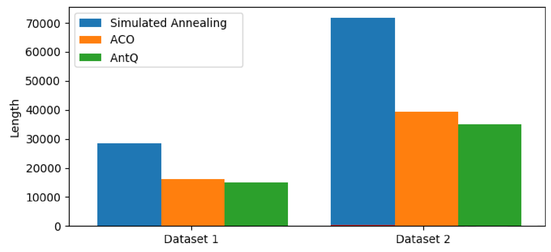

Chart 2. Best results produced by Ant-Q, ACO and SA algorithm

Experiment Result

Chart 1 indicating the result produced by 7 trials. With the solutions lying between 14446 and 14943. The algorithm has proved that it is consistent enough to produce decent results. It is observable that the solution for this specific problem is found when the iteration reach around the number of 700.

In chart 2, we compared the average behavior of Ant-Q with following well-known heuristic algorithms: Simulated Annealing(SA) and Ant Colony Optimization(ACO). The obtained result is show that Ant-Q was almost always the best performing algorithm.

Future Plans

Our system is built for the only purpose that solving the delivery problem. Not only researching, testing, collecting data to make a conclusion, a comparison or a statistic. Moreover we want to apply the solution to the practical application.

About academic researching:

- Our system allows us scale out our model, we also can apply the other algorithms which belong to other concepts of machine learning.

About practical application:

- Contributing to build solutions for Logistic.
- Optimizing existing navigation algorithms.
- Being a tool for developing Artificial intelligence (AI) which is specifically the automatic navigation robot.

Conclusions

Our system is built for the only purpose that solving the delivery problem. Not only researching, testing, collecting data to make a conclusion, a comparison or a statistic. Moreover we want to apply the solution to the practical application.

References

[1] Colorni A., M. Dorigo and V. Maniezzo, 1991. Distributed Optimization by Ant Colonies. Proceedings of ECAL91 - European Conference on Artificial Life, Paris, France, F.Varela and P.Bourgine (Eds.), Elsevier Publishing, 134–142.
[2] Colorni A., M. Dorigo and V. Maniezzo, 1992. An Investigation of some Properties of an Ant Algorithm. Proceedings of the Parallel Problem Solving from Nature Conference (PPSN 92), Brussels, Belgium, R.Männer and B.Manderick (Eds.), Elsevier Publishing, 509–520.
[3] Gambardella L. and M. Dorigo, 1995. Ant-Q: A Reinforcement Learning approach to the traveling salesman problem. Proceedings of ML-95, Twelfth International Conference on Machine Learning, Tahoe City, CA, A. Prieditis and S. Russell (Eds.), Morgan Kaufmann, 252–260.
[4] Péter Stefán, László Monostori, Ferenc Erdélyi. Reinforcement learning for solving shortest-path and dynamic scheduling problems.
[5] Csaba Szepesvásri, 2009. Algorithms for Reinforcement Learning. Draft of the lecture published in the Synthesis Lectures on Artificial Intelligence and Machine Learning series by Morgan & Claypool Publishers
[6] David Arthur, Sergei Vassilvitskii. k-means++: The Advantages of Careful Seeding