

강화학습 기반의 차량 경로 문제 일반화 방안 연구

정철환[○], 김한솔*, 김광수*

[○]성균관대학교 인공지능학과

*성균관대학교 소프트웨어학과

e-mail : jeriks@g.skku.edu[○], terry5969@g.skku.edu*, kim.kwangsus@skku.edu*

A Study of Solving the Generalized Vehicle Routing Problem Using Reinforcement Learning

Chul-Hwan Jung[○], Han-Sol Kim*, Kwang-Su Kim*

[○]Dept. of Artificial Intelligence, Sungkyunkwan University

*College of Computing and Informatics, Sungkyunkwan University

요 약

본 연구에서는 기존 차량 경로 문제(Vehicle Routing Problem)의 범위를 확장시켜 일반화된 차량 경로 문제(Generalized Vehicle Routing Problem)를 제시하고, 이 문제를 해결하기 위한 강화학습 모델을 제안한다. 기존의 차량 경로 문제는 depot에서 각 node(또는 각 node에서 depot)의 단방향만 존재해 제한된 문제만을 해결할 수 있었다. 이 한계점을 극복하기 위해 depot을 제외한 모든 node가 서로 연결된 형태의 일반화된 차량 경로 문제를 정의하고 이를 해결하고자 한다. 차량 경로 문제는 NP-hard 문제로 최근에는 강화학습을 이용해 이를 해결하고자 하는 모델이 연구되고 있다. 본 연구에서는 새로 정의한 일반화된 차량 경로 문제를 해결하기 위한 강화학습 모델을 제안한다.

- ▶ Keyword : 일반화된 차량 경로 문제(generalized vehicle routing problem), 조합 최적화(combinatorial optimization), 강화학습(reinforcement learning)

1. Introduction

차량 경로 문제(Vehicle Routing Problem, 이하 'VRP')는 출발지(depot)에서 출발한 차량이 배달이나 수집 등의 서비스를 요구하는 고객(node)을 방문하고, 다시 출발지로 돌아오는 최소 비용의 경로를 결정하는 문제이다[1]. VRP는 실제 물류 산업에서 비용 절감의 큰 역할을 하고 있다. 물류센터에서 고객에게 배송하는 차량의 최적 경로를 VRP를 통해 계산할 수 있기 때문이다. 그러나 VRP는 Fig. 1. (a)와 같이 depot에서 node(또는 node에서 depot)의 단방향의 경로만을 고려하여 정의된 문제로, Fig. 1. (b)와 같이 양방향의 경로가 있는 문제에는 적용할 수 없는 한계점이 있다. 본 연구에서는 이를 해결하기

위해 depot을 제외하고 node와 node 사이에 양방향의 경로가 있는 문제를 일반화된 차량 경로 문제(Generalized Vehicle Routing Problem, 이하 'GVRP')라 정의한다.

VRP는 NP-hard 문제로 node의 개수가 증가함에 따라 경우의 수가 무한히 늘어나게 되어 알고리즘으로 최적 경로를 찾는 데 한계가 있는 문제이다. 이 한계점을 극복하기 위해 최근엔 강화학습을 이용해 최적 경로를 찾는 연구가 활발히 이루어지고 있다[2]. GVRP는 VRP를 일반화한 모델로써, node의 개수가 동일할 경우 훨씬 더 많은 경로의 경우의 수가 생기게 된다. 이를 해결하기 위해 본 연구에서는 새로 정의된 GVRP의 최적 경로를 찾기 위한 강화학습 모델을 제안한다.

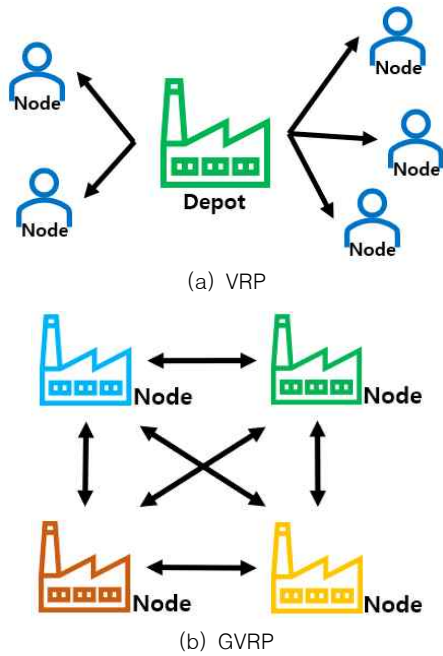


Fig. 1. VRP and GVRP

II. Preliminaries

1. Related works

1.1 외판원 문제(Traveling Salesman Problem)

외판원 문제(Traveling Salesman Problem, 이하 ‘TSP’)는 여러 도시들이 있고 한 도시에서 다른 도시로 이동하는 비용이 주어졌을 때, 모든 도시들을 한번만 방문하고 원래 시작점으로 돌아오는 최소 비용의 이동 순서를 찾는 문제이다. TSP는 조합 최적화의 대표적인 문제로 최적 알고리즘과 휴리스틱 알고리즘을 이용한 연구가 많이 있었다[4].

그러나 TSP의 node의 개수가 100개 이상이 되면 기존 알고리즘으로 최적 경로를 찾는 데 한계가 있어, 최근에는 딥러닝 및 강화학습을 이용해 최적 경로를 찾는 모델이 제안되고 있다. 대표적으로 [6]의 Pointer Network 모델은 각 node의 좌표값을 입력으로 받아 attention mechanism을 이용해 node 간의 관계를 학습하고, 지도학습을 이용해 node의 경로 순서(index)를 출력으로 하여 최적 경로를 찾는 모델을 제안했다. [7]에서는 기존 Pointer Network의 지도학습 모델은 학습 데이터 이상의 최적 경로를 찾을 수 없다는 한계점을 지적하고, 최적 경로의 길이를 최소화하는 방법으로 학습하는 강화학습을 모델을 제안했다.

1.2 차량 경로 문제(Vehicle Routing Problem)

VRP는 TSP의 일반화된 문제로, 1959년 Dantzig에

의해 처음 제기되어 수십년 동안 응용 수학과 컴퓨터 과학에서 연구되어 온 조합 최적화 문제이다[3]. 최적 알고리즘과 휴리스틱 알고리즘을 이용해 많은 연구가 이루어졌지만, 기존 알고리즘으로는 node의 수가 증가함에 따라 최적의 해를 찾는 데 한계를 보이고 있다. 이를 해결하기 위해 최근에는 강화학습을 이용한 모델이 연구되고 있다.

[8]은 [6]에서 제안한 Pointer Network를 개선한 모델을 제안했다. Pointer Network의 RNN 인코더를 element-wise projection으로 대체하여 업데이트된 node의 임베딩 정보를 효과적으로 계산할 수 있다. node 임베딩 정보로 각 node의 관계를 학습할 수 있고, 학습된 node간의 관계를 이용해 최적의 경로를 찾는 모델을 제안하였다. [9]는 인코더-디코더 구조로, 인코더에서는 attention layer를 이용해 node와 node 사이의 관계를 학습하게 된다. 디코더에서는 학습된 node의 관계를 이용하여 최적 경로를 찾는 모델을 제안했다.

III. The Proposed Scheme

본 연구에서는 용량제한이 있는 차량 경로 문제(Capacitated Vehicle Routing Problem, 이하 ‘CVRP’)를 다룬다. CVRP는 VRP와 같은 depot-node 구조에 depot에서 node(또는 node에서 depot)로 배송해야 하는 요구 수량이 있고, 배송하는 차량의 용량제한이 있어 용량만큼의 수량만 배송할 수 있다.

강화학습을 이용한 CVRP의 최적 경로를 찾는 과정은 먼저 node의 정보를 이용해 node와 node 사이의 관계를 학습하고, 학습된 관계 정보를 이용해 강화학습 모델로 최적 경로를 찾게 된다. 이때, node와 node 사이의 관계를 학습하기 위해서는 node의 특성을 모두 반영하여 임베딩 시키는 과정이 필요하다.

CVRP에서 node의 특성 값은 좌표값과 요구 수량이 된다. n 개의 node의 집합 $A=\{a_i | i=1,2,...,n\}$ 에서 임베딩 입력 특성 값은 node의 좌표값(a_{ix_i} , a_{iy_i})과 depot의 요구 수량 a_{id_i} 이 된다. 그러나 GVRP에서는 node와 node가 모두 연결된 문제로 연결된 모든 경로에는 요구 수량이 존재한다. 모든 경로의 요구 수량은 2차원의 형태로 표현이 되고, 각 node의 요구 수량은 집합 $a_{iD} = \{a_{id_i} | i=1,2,...,n\}$ 으로 표현할 수 있다. 따라서 GVRP에서 각 node의 임베딩 입력 특성 값은 좌표값(a_{ix_i} , a_{iy_i})과 요구 수량의 집합 a_{iD} 이 된다.

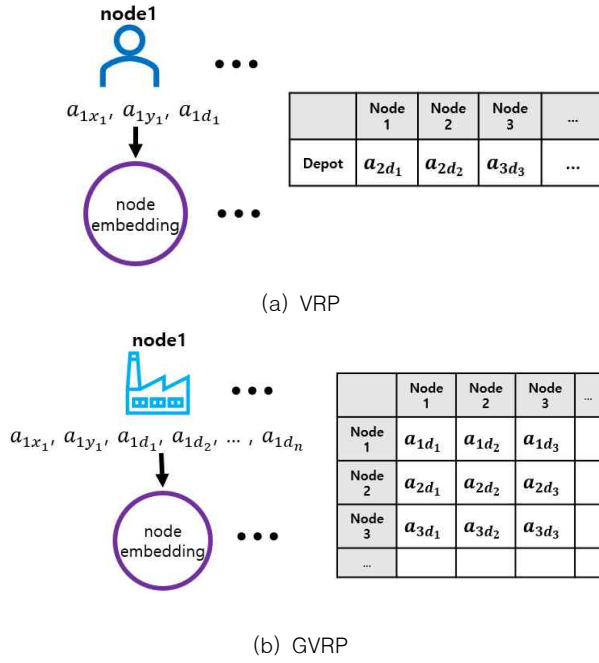
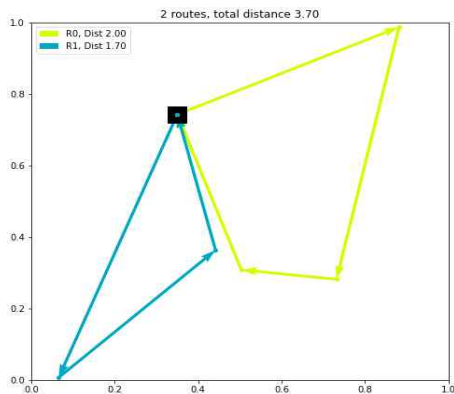


Fig. 2. Node embedding

각 node의 정보를 반영하여 생성된 node 임베딩으로 node 간의 관계를 학습할 수 있고 학습된 node 간의 관계를 이용해 강화학습 모델로 최적 경로를 찾을 수 있다. 강화학습은 exploration과 exploitation을 통해 차량 경로를 최소화 시키는 최적 경로를 찾게 된다. 강화학습 모델의 결과를 보면 VRP의 경우 depot에서 출발하여 요구 수량이 있는 node를 차량 용량만큼 배달 후 다시 depot으로 돌아온다. GVRP에서는 한 방향이 아닌 node와 node 간의 요구 수량을 모두 처리하기 위해 복잡한 이동 경로가 생성된다.



(a) VRP(n=5)

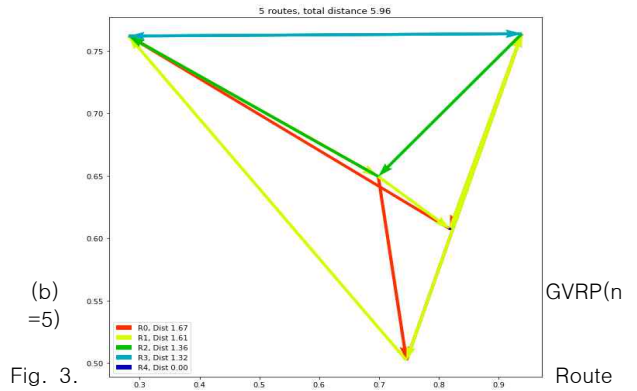


Fig. 3.

IV. Conclusions

본 연구에서는 VRP의 일반화된 문제인 GVRP를 제안함으로써, 모든 node가 연결된 구조의 차량 경로 문제를 정의할 수 있었다. 또한, 새로 정의한 GVRP에서 node의 추가된 정보를 임베딩 시킬 수 있는 방법을 제안함으로써 강화학습을 이용해 최적 경로를 찾을 수 있는 모델을 제안했다.

본 논문의 모의실험에서는 VRP를 일반화 시킨 GVRP 문제에서도 강화학습을 이용하여 각 node의 요구 수량을 변경하고 node와 node 간의 요구 수량을 모두 처리하는 동시에 경로를 최소화 시키는 최적의 경로를 찾을 수 있었다.

모의실험 결과, 본 연구에서 제안한 VRP를 일반화 시킨 강화학습 기반의 GVRP 모델을 통해 GVRP 및 VRP의 문제에서 최적 경로를 찾는 것이 가능할 것으로 기대된다.

Acknowledgement

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No.2020-0-00973, VR·AR 콘텐츠 비가시 영역 영상 복원 기술 개발)과 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-02068, 인공지능 혁신 허브 연구 개발)

References

- [1] Paolo Toth, "Vehicle routing: problems, methods, and applications." SIAM, 2014.
- [2] A Bogrybayeva, "Learning to Solve Vehicle Routing Problems: A Survey" arXiv:2205.02453, 2022.
- [3] Dantzig G., "The Truck Dispatching Problem."

Management Science, pp. 80–91, 1959.

- [4] Daniel J Rosenkrantz, "An analysis of several heuristics for the traveling salesman problem." Fundamental Problems in Computing, pp. 45–69., 2009.
- [5] Michel Deudon, "Learning heuristics for the TSP by policy gradient" In International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research, pp. 170–181, 2018.
- [6] Oriol Vinyals, "Pointer networks." In Advances in Neural Information Processing Systems, pp. 2692–2700, 2015.
- [7] Irwan Bello, "Neural combinatorial optimization with reinforcement learning" arXiv:1611.09940, 2016.
- [8] MohammadReza Nazari, "Reinforcement learning for solving the vehicle routing problem." In Advances in Neural Information Processing Systems, pp. 9860–9870, 2018.
- [9] Wouter Kool, "Attention, Learn to Solve Routing Problems!" International Conference on Learning Representations, 2019.