https://doi.org/10.7840/kics.2025.50.10.1515

동적 환경에서의 차량/화물 최적 배차를 위한 MAML과 PPO 통합 알고리즘

김 지 현•, 권 순 영*, 신 다 민*, 김 형 남°

Integration of MAML and PPO for Optimized Vehicle/Cargo Dispatch in Dynamic Environments

Ji-Hyeon Kim•, Soon-Young Kwon*, Da-Min Shin*, Hyoung-Nam Kim°

요 약

항만물류 환경에서 차량과 화물 간의 효율적인 배차를 위해서는 배차의 공정성/일관성 및 도로 상황 등과 같은 다양한 환경 요소를 고려해야 한다. 이러한 요소들은 실시간으로 변화하며, 운송 차량과 운전자에게 직접적인 영향을 미칠 수 있다. 하지만 기존의 최적화 기법은 환경 변화에 대한 적응력이 부족하여 실시간 배차 최적화에는 한계가 존재한다. 본 논문에서는 메타 강화학습 알고리즘인 MAML(model-agnostic meta-learning)과 PPO(proximal policy optimization)를 결합한 배차 최적화 기법을 제안한다. MAML을 활용하여 다양한 배차 환경에서 최적화된 초기 정책을 학습하고, PPO를 통해 환경 변화에 적응하며 지속적으로 정책을 개선함으로써 배차 성능을 극대화한다. 모의실험을 통해 MAML+PPO 기반 최적화 알고리즘이 기존 알고리즘 대비 우수한 배차 성능을 보임을 확인하였다. 특히, APM(aggregate performance measure) 기반 성능 분석 결과, 제안한 알고리즘이 다양한 시나리오에서 높은 일반화 성능과 안정적인 최적화 성능을 달성함을 보였다.

키워드 : 차량/화물 배차 시스템, 운송 매칭 알고리즘, MAML, PPO
Key Words : Vehicle/cargo dispatch system, Transportation matching algorithm, MAML, PPO

ABSTRACT

For efficient vehicle and cargo dispatch in port logistics environments, it is essential to consider various environmental factors such as dispatch fairness/consistency, road conditions. These factors change dynamically in real-time and can significantly impact transportation vehicles and drivers. However, conventional optimization methods struggle to adapt to these environmental changes, making real-time dispatch optimization challenging. In this paper, we propose a dispatch optimization approach that integrates model-agnostic meta-learning(MAML) and proximal policy optimization(PPO). MAML is employed to learn an optimized initial policy across diverse dispatch environments, and PPO is used to adapt to environmental changes and continuously refine the policy, thereby maximizing dispatch performance. Through simulations, we demonstrate that the MAML+PPO-based optimization algorithm outperforms traditional methods in dispatch performance. In particular, the aggregate

※ 이 논문은 2023년도 정부(중소벤처기업부)의 재원으로 중소기업기술정보진흥원의 지원을 받아 수행된 연구임 (NTIS 고유번호 : 1425178206).
• First Author : Department of Electrical and Electronics Engineering, Pusan National University, kjihyeon@pusan.ac.kr, 학생회원
° Corresponding Author : School of Electrical and Electronics Engineering, Pusan National University, hnkim@pusan.ac.kr, 종신회원
* Department of Electrical and Electronics Engineering, Pusan National University, {ysk1680, ekals2020}@pusan.ac.kr, 학생회원
논문번호 : 202503-054-A-RE, Received March 10, 2025; Revised April 15, 2025; Accepted April 29, 2025

performance measure(APM)-based performance analysis confirms that the proposed algorithm achieves high generalization performance and stable optimization across various scenarios, making it a robust solution for adaptive dispatch optimization in dynamic environments.

Ⅰ. 서 론

항만물류 환경에서 차량/화물 배차 기술은 운송 수요 증가와 복잡한 물류 운영을 효과적으로 처리하기 위해 필수적인 요소로 자리 잡고 있다. 현재 국내 항만 물류 시스템은 디지털 전환 속도가 더디고, 기존의 수작업 중심의 배차 방식으로 인해 비효율성이 발생하고 있다. 특히, 협력 운송사 간에 배차 협의를 진행할 수 있는 체계적인 시스템이 부족하여 불공정한 배차가 이루어지거나, 업무원의 임의적인 배차 선정에 의존하는 경향이 있어 배차의 공정성을 보장하기 어렵다. 또한, 운송 지연 문제는 차량과 화물의 배차 최적화가 부족한 상황에서 더욱 심화되고 있다. 이를 해결하기 위해 항만물류 시스템의 고도화 및 디지털 전환을 통한 배차 자동화 연구에 대한 수요가 증가하고 있으며[1,2], 강화학습(reinforcement learning)을 활용한 운송 최적화 연구가 활발히 진행되고 있다[3,4].

일반적으로 차량/화물 배차는 미배차 화물, 차량 가용성, 화물의 출발지 및 목적지 정보 등을 기반으로 이루어진다. 하지만, 교통 혼잡도, 날씨, 차량 상태, 운송 거리 등 실시간으로 변화하는 환경적 요인은 배차 성능에 중요한 영향을 미친다. 이러한 요인을 고려하지 않은 배차 시스템은 최적의 운송 효율을 달성하기 어려우며, 결과적으로 불균형한 배차와 운송 지연을 초래할 수밖에 없다.

전통적인 최적화 기법으로는 유전 알고리즘(genetic algorithm, GA)[5]과 규칙 기반 휴리스틱(rule-based heuristic, RBH)[6]이 있다. GA는 여러 배차 방법을 생성하고 그중 성능이 좋은 방법을 선택해 점진적으로 개선하는 방식으로, 다양한 조합을 고려할 수 있지만 높은 연산 비용과 긴 수렴 시간이 요구된다. 반면, RBH는 특정 규칙을 기반으로 차량과 화물을 배차하는 방식으로 계산 속도는 빠르지만, 환경 변화에 적응하지 못해 복잡한 운송 환경에서는 최적의 배차 성능을 보장하기 어렵다. 이러한 기존 최적화 기법들은 동적이고 복잡한 운송 환경에서 충분한 적응성을 제공하지 못한다는 단점이 있다.

강화학습이 이러한 문제점을 해결하기 위한 대안으로 적용되고 있으며, 특히 다양한 운송 환경에 빠르게 적응할 수 있는 메타 강화학습(meta reinforcement learning)이 주목받고 있다[7-9]. 메타 학습은, 적은 양의 데이터로도 새로운 환경에 모델이 빠르게 적응할 수 있도록 하는 기법이다[10]. 대표적인 기법으로는 최적화 기반 메타 강화학습 알고리즘인 MAML(model-agnostic meta-learning)[11,12]이 있으며, 다양한 환경에서도 적은 학습만으로 효과적인 정책을 생성할 수 있도록 최적화된 초기 정책을 제공하는 데 강점을 가진다. MAML을 이용하면 다양한 배차 시나리오에서 빠르게 적응할 수 있으며, 기존의 강화학습 알고리즘보다 적은 학습 단계만으로도 높은 성능을 달성할 수 있다. 하지만 MAML은 초기 정책 최적화에는 효과적이지만, 지속적인 정책 개선을 수행하기에는 한계가 있다. 이를 극복하기 위해서는 정책을 점진적으로 최적화할 수 있는 기법이 필요하다.

PPO(proximal policy optimization)는 정책 기반 강화학습 기법으로, 정책 업데이트 과정에서 클리핑(clipping) 기법을 활용하여 안정적인 학습을 보장한다[13]. 이를 통해 환경 변화에 적응하면서도 장기적인 최적화를 수행하는 데 강점을 가진다.

본 연구에서는 MAML을 활용하여 초기 정책을 학습하고, PPO를 이용하여 지속적인 정책 개선을 수행함으로써 배차 성능을 최적화하는 방식을 적용하기 위해 MAML과 PPO를 결합한 새로운 최적화 알고리즘을 제안한다. 이를 통해 다양한 운송 환경에서 높은 일반화 성능을 갖춘 배차 시스템을 개발하고자 한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 차량/화물 배차 시스템 모델을 설명하고, Ⅲ장에서는 본 연구와 관련된 기존 연구들을 살펴보며 한계점을 분석한다. Ⅳ장에서 기존 연구의 한계를 보완하기 위해 MAML과 PPO를 결합한 최적화 알고리즘을 제안한다. Ⅴ장에서는 모의실험을 통해 제안한 알고리즘의 성능을 분석하고, Ⅵ장에서 본 논문의 결론을 맺는다.

Ⅱ. 차량/화물 배차 시스템 모델

항만물류 환경에서 차량과 화물의 배차는 운송 수요 증가와 복잡한 물류 운영을 효과적으로 처리하기 위해 필수적인 요소로 자리 잡고 있다. 기존의 배차 시스템은 주로 예약된 화물을 운송 가능한 차량에 수작업으로 배정하는 방식으로 운영되었으며, 이는 배차 공정성 저하

와 운송 효율성 감소를 초래하는 주요 원인이 된다. 또한, 차량과 화물의 배차 과정에서 운송 거리, 교통 상황, 운전자의 운행 시간 제한 등의 변수들이 고려되지 않거나 제한적으로 반영되어 최적의 운송 성능을 달성하기 어렵다. 이를 해결하기 위해, 본 장에서는 다양한 운송 환경에서의 강화학습 기반 배차 최적화를 위한 차량/화물 배차 시스템 모델을 설계한다.

그림 1은 N 개의 화물과 이를 운반할 M 대의 트럭이 있다고 가정하였을 때의 차량/화물 배차 시스템 모델을 나타낸 것이다. 차량과 화물의 최적 배차를 위해 거리 기반, 시간 기반, 교통 상황 기반의 다양한 배차 시나리오를 정의한다.

먼저, 거리 기반 배차는 단거리, 중거리, 장거리 운송으로 구분되며, 각 운송 거리별 최적화 목표가 다르게 설정된다. 단거리 운송에서는 배차 빈도를 높이고 신속한 배차를 수행하는 것이 중요하며, 장거리 운송에서는 차량 활용률과 적재량을 최적화하는 것이 핵심 요소가 된다. 중거리 운송의 경우, 이동 시간과 화물 적재량을 균형 있게 조절하는 것이 중요한 요소로 작용한다.

시간 기반 배차는 주간과 야간으로 구분되며, 각 시간대에 따른 배차 전략이 다르게 적용된다. 주간 운송의 경우, 교통 혼잡이 심하여 운송 지연을 최소화하는 방향으로 최적화가 이루어져야 한다. 반면, 야간 운송은 상대적으로 원활한 교통 흐름을 활용하여 차량 활용도를 높이는 전략이 필요하다.

교통 상황 기반 배차에서는 교통 혼잡도에 따라 배차 알고리즘이 다르게 작동하도록 설계되었다. 혼잡 구간에서는 최적의 경로를 탐색하여 운송 지연을 줄이고, 원활한 도로 환경에서는 배차의 균형성과 차량 운영 효율을 극대화하는 방식이 적용된다.

이러한 시나리오는 실제 항만물류 운송 환경을 반영하여 배차 알고리즘의 성능을 평가하는 기준으로 활용된다.

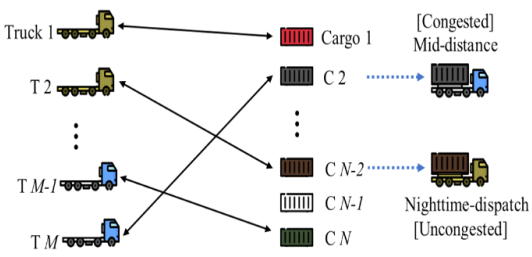


그림 1. 차량/화물 배차 시스템 모델
Fig. 1. The vehicle/cargo dispatch system model.

Ⅲ. 관련 연구

3.1 전통적 배차 최적화 기법

전통적인 차량/화물 배차 최적화 기법은 대표적으로 유전 알고리즘(genetic algorithm, GA)과 규칙 기반 휴리스틱(rule-based heuristic, RBH)이 있다. 유전 알고리즘은 진화 이론에서 영감을 받은 집단 기반 탐색 방법으로, 초기 해 집단을 무작위로 생성한 후 선택, 교차(crossover), 변이(mutation) 연산을 반복하여 점진적으로 최적 해에 접근하는 방식이다[5]. GA는 탐색 공간을 효과적으로 넓히고 다양한 후보 해를 탐색할 수 있어 글로벌 최적 해에 근접할 가능성이 높지만, 많은 후보 해를 탐색하는 과정에서 높은 연산 비용과 긴 수렴 시간을 필요로 한다는 단점이 있다. 따라서 실시간 변화가 많은 동적 배차 환경에서는 효율성이 저하될 수 있다.

RBH는 명확하고 직관적인 규칙을 사용하여 차량과 화물을 배치하는 접근 방법이다. 일반적으로 차량의 위치, 차량 상태, 화물의 위치와 배송 기한 등의 정해진 규칙에 따라 빠르게 배차 결정을 내릴 수 있어 연산 속도가 매우 빠르고 효율적이다[6]. 그러나 이 방법은 배차 결정의 성능이 규칙의 설계 품질에 크게 의존하며, 사전에 예측하지 못한 교통 혼잡, 차량 고장, 날씨 변화와 같은 돌발 상황에 효과적으로 대응하기 어렵다. 또한, 다양한 조건이나 복잡한 배차 시나리오를 반영하려면 규칙이 복잡해지고 관리가 어려워지는 단점이 존재한다.

3.2 심층 강화학습 기반 배차 최적화[3,4]

심층 강화학습(deep reinforcement learning, DRL)은 기존 배차 최적화 방식의 단점을 개선하고, 실시간으로 변화하는 환경에 적응하는 배차 문제 해결을 위한 접근 방법으로 주목받고 있다. DRL은 환경과의 상호작용을 통해 축적된 경험 데이터를 활용하여 정책을 최적화하는 방법론으로, 운송 환경의 복잡성과 실시간 변동성을 효과적으로 처리할 수 있다[3,4]. DRL 기반 방법들은 환경 변화에 대한 높은 적응력을 보이며, 특히 신경 조합 최적화(neural combinatorial optimization) 기법과 결합하여 차량 경로 문제(vehicle routing problem)와 같은 복잡한 물류 문제에서 뛰어난 성능을 발휘하고 있다. 그러나 DRL 접근법은 일반적으로 많은 데이터와 긴 초기 학습 시간이 요구되며, 정책 초기 설정이 부적절할 경우 성능이 불안정해질 수 있다는 한계를 지닌다. 또한, 특정 환경에 과적합(overfitting)될 가능성도 있어 다양한 배차 환경에 대한 일반화된 성능 확보가 쉽지 않다.

3.3 메타 강화학습과 MAML 및 PPO 기반 최적화 접근법

메타 학습(meta learning)은 다양한 환경에서 얻은 학습 경험을 일반화하여, 새로운 환경에서도 빠르게 적응할 수 있도록 하는 학습 기법이다[7-9]. 특히 메타 강화학습(meta reinforcement learning)은 메타 학습의 일종으로, 강화학습의 프레임워크를 이용하여 환경 변화에 빠르게 적응할 수 있는 정책을 생성한다[10].

MAML(model-agnostic meta-learning)은 대표적인 메타 강화학습 방법으로, 다양한 작업(task)에서 얻은 경험을 바탕으로 소량의 데이터만으로도 신속하게 초기 정책을 최적화할 수 있도록 한다[11,12]. 그러나 MAML은 초기 정책 학습에 뛰어나지만, 지속적인 환경 변화에 따른 장기적인 정책 개선 능력은 제한적이다.

PPO(proximal policy optimization)는 정책 업데이트 시 급격한 변화를 제한하는 클리핑(clipping) 기법을 활용하여 안정적이고 점진적으로 정책을 최적화할 수 있는 강화학습 기법이다[13]. 기존의 정책 기반 강화학습 알고리즘인 TRPO(trust region policy optimization)[14]는 정책 업데이트 시 변화를 제한하는 방식으로 학습을 안정화할 수 있지만, 계산 비용이 높다는 단점이 있다. 반면, PPO는 상대적으로 간단한 클리핑 기법을 활용하여 정책의 급격한 변화를 방지하며 환경 변화에 안정적으로 대응할 수 있어 장기적인 정책 개선에 강점을 가진다. 다만, PPO의 성능은 초기 정책 설정에 따라 편차가 크게 나타날 수 있으며, 이로 인해 초기 단계에서 효율성이 저하되는 문제가 발생한다.

최근 연구에서는 이러한 MAML과 PPO의 장점을 결합하여, 초기 정책을 빠르게 설정하고 이후 정책을 지속적으로 개선하는 통합 알고리즘 연구들이 진행되고 있다. 로봇 제어 분야에서는 다양한 작업 환경에 빠르게 적응할 수 있는 MAML 기반 초기 정책을 PPO를 통해 지속적으로 개선하여 높은 성능과 강건성을 입증하였으며[15], 이동통신 네트워크 분야에서도 트래픽 변화에 신속히 대응하고 자원 관리 효율성을 향상시키기 위해 메타 학습과 PPO를 결합한 접근법을 적용하였다[16]. 그러나 복잡하고 동적인 차량 및 화물 배차 환경을 대상으로 한 적용 연구는 여전히 제한적이다.

본 논문에서는 이러한 기존 연구들의 한계를 보완하고자, MAML을 통해 우수한 초기 정책을 학습하고 PPO를 활용하여 지속적인 정책 최적화를 수행하는 차량/화물 배차 통합 알고리즘을 제안하여 다양한 운송 환경에서 성능을 검증하고자 한다.

Ⅳ. MAML과 PPO를 결합한 최적화 알고리즘

Ⅲ장 관련 연구에서 제시한 기존의 최적화 방법은 복잡한 환경 변화에 적응하거나 배차 효율성 측면에서 한계가 존재한다.

따라서 본 장에서는 항만물류 환경의 차량/화물 배차 문제에 맞추어 MAML을 활용하여 차량/화물 배차의 초기 정책을 신속하게 최적화하고, 이후 PPO를 적용하여 지속적인 정책 개선을 수행하는 최적화 기법을 제안한다.

4.1 MAML(model-agnostic meta-learning) 기반 초기 정책 학습[11,12]

MAML 기반 메타 강화학습에서는 그림 2와 같이 배차 문제를 해결하기 위해 다수의 배차 시나리오를 task로 정의하고, 이를 활용하여 초기 정책을 최적화하는 과정을 수행한다. 먼저, 다양한 운송 환경을 반영한 배차 시나리오 $\mathcal{T} = \{T_1, T_2, \cdots, T_N\}$를 샘플링하여 메타 학습에 활용한다. 각 task T_i는 단거리, 중거리, 장거리 운송과 같은 거리 기반 시나리오, 혼잡도 수준, 운송 시간대 등의 다양한 환경 요인을 포함하도록 구성된다.

각 task T_i에 대하여, 기존의 정책 θ를 사용하여 손실 함수 $L_{T_i}(\theta)$를 계산하고, 경사하강법(gradient descent)을 통해 식 (1)과 같이 θ'로 파라미터 업데이트를 수행한다.

$$\theta_i' = \theta - \alpha \nabla_\theta L_{T_i}(\theta) \tag{1}$$

여기서 α는 학습률을 나타내며 task별 손실 함수 $L_{T_i}(\theta)$에 대한 그래디언트 $\nabla_\theta L_{T_i}(\theta)$를 기반으로 최적화가 수행된다. 이렇게 task별 학습이 이루어진 후, MAML의 핵심 과정인 메타 업데이트(meta-update)가

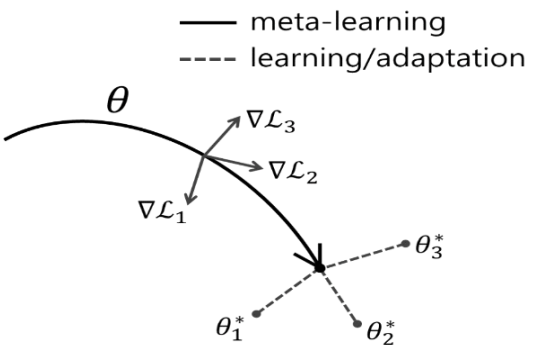


그림 2. MAML의 최적화 과정[11]
Fig. 2. MAML optimization process.

수행된다. 메타 업데이트 단계에서는 여러 task에서 학습된 정책을 기반으로, 전체 task 집합에 대해 초기 정책 θ를 최적화하는 과정이 진행된다. 이를 통해 다양한 task에서 빠르게 적응할 수 있는 일반화된 정책을 도출할 수 있으며, 메타 업데이트는 다음과 같이 정의된다.

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{T_i \in \mathcal{T}} L_{T_i}(\theta_i'), \quad (2)$$

여기서 β는 메타 학습률을 나타내며 여러 task에서의 손실을 최소화하는 방향으로 정책을 업데이트한다. 이러한 과정이 반복되면서, 새로운 배차 환경에서도 적은 학습 단계만으로 높은 성능을 달성할 수 있는 초기 정책이 학습된다.

4.2 PPO(proximal policy optimization) 기반 정책 최적화[13]

PPO 기반 정책 학습에서는 배차 시스템의 상태 s_t와 이에 대응하는 행동 a_t를 입력으로 받아, 각 행동에 대한 정책 확률을 $\pi_{\theta}(a_t|s_t)$로 모델링한다. 정책 업데이트 과정에서 PPO는 식 (3)과 같은 목적함수를 최적화한다.

$$L^{CLIP}(\theta) = \mathbb{E}[\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)], \quad (3)$$

여기서 $r_t(\theta)$는 정책의 업데이트 비율을 나타내며 현재 정책 $\pi_{\theta}(a_t|s_t)$와 이전 정책 $\pi_{\theta_{old}}(a_t|s_t)$ 의 비율로 계산된다. A_t는 어드밴티지 함수(advantage function)로, 특정 행동 a_t가 얼마나 좋은 선택이었는지를 나타낸다. ϵ은 클리핑 파라미터로, 일반적으로 작은 값 (0.1~0.2)를 설정하여 정책이 급격하게 변화하는 것을 방지한다.

이 목적 함수는 정책 업데이트 시 클리핑 기법을 적용하여 정책의 급격한 변화를 제한하며, 이를 통해 학습의 안정성을 유지한다. 또한, 배차 시스템의 최적화 목표에 맞춰 보상 함수 R_t를 정의하여 모델의 배차 성능을 극대화할 수 있도록 한다.

$$R_t = w_1 \times E_t + w_2 \times U_t + w_3 \times C_t + w_4 \times D_t, \quad (4)$$

여기서 E_t는 배차 완료까지 소요된 평균 시간으로 지연 시간 감소율을 반영하며, U_t는 차량의 총 운행 시간 대비 실질적인 운송 시간 비율을 의미한다. 그리고, C_t는 단위 시간 내에 처리된 화물량이며, D_t는 특정 차량에 배차가 집중되지 않도록 하는 공정성 지표를 나타낸다. 각 성능 지표의 상대적 중요도는 가중치 w_1, w_2, w_3, w_4를 통해 조정되며, 총합이 1이 되도록 설정한다. 가중치의 값들은 배차 시스템의 운영 목표에 따라 달라질 수 있는데, 예를 들어, 단거리 운송에서는 시간 효율성이 중요하므로 w_1 값을 높게 설정할 수 있다. 또한, 장거리 운송에서는 한 번의 운송에서 많은 화물을 처리하는 것이 중요하므로 w_3 값을 상대적으로 높게 설정할 수 있다.

그러나, R_t는 특정 시점의 배차 성능만을 반영하므로, 전체 운영 기간 동안의 정책 성능을 평가하기에는 한계가 있다. 따라서, 본 논문에서는 APM(aggregate performance measure)을 도입하여, 정책이 다양한 환경에서 일관된 성능을 유지하는지를 평가한다[17].

$$APM = \frac{1}{T}\sum_{t=1}^{T} R_t, \quad (5)$$

여기서 T는 전체 운영 시간(혹은 학습된 에피소드 수)이며, R_t는 각 시점에서의 보상을 의미한다. APM은 단일 시점이 아닌 장기적인 배차 성능을 반영하는 지표로 활용되며, 학습 과정에서 정책이 안정적으로 수렴하는지를 확인하는 데 사용된다.

Ⅴ. 차량/화물 배차 모의실험

본 장에서는 차량/화물 배차 시스템에서 다양한 환경 요소에 따른 MAML+PPO 결합 알고리즘 기반 운송 매칭 기법의 배차 성능을 검증하기 위한 모의실험을 진행한다. 배차 성능은 시간 효율성(time efficiency, TE), 차량 활용도(vehicle utilization, VU), 화물 처리량(cargo throughput, CT), 배차 균등성(dispatch equity, DE), 총괄 성능 지표(aggregate performance measure, APM) 등을 통해 평가된다.

5.1 실험 환경 설정

Ⅱ장에서 설명한 배차 시스템 모델과 동일하게, 각 화물과 차량에는 운송 거리 및 운송 시간과 같은 임의의 정보를 부여한다. 또한, 배차 과정에서는 운송 거리, 교통 체증 등 다양한 운송 환경 요소가 반영된다. 구체적인 모의실험 파라미터는 표 1과 같다. 장거리 운송 가능 거리는 인천화물터미널에서 부산까지의 거리를 고려하여 400 km로 가정하였다. 차량 속도는 화물차의 법정 최고속도인 90 km/h를 기준으로 하되, 빙판길 주행 시 최고속도의 50%로 감속 제한이 적용됨을 반영하여 최소 30 km/h, 최대 80 km/h 범위로 설정하였다. 학습 및 시험 과정은 Intel(R) Core(TM) i5-8600K CPU 환경에서 수행되었으며, Pytorch 프로그램[18]을 활용하여

표 1. 모의실험 환경 파라미터
Table 1. Simulation environment parameters.

Parameters	Values
Maximum transport distance	400 km
Transportable distance (short/mid/long)	50 km/200 km/400 km
Maximum driving time	2~6 hours
Traffic congestion	congested/uncongested
Vehicle speed range	30~80 km/h

알고리즘을 구현하였다. 학습 단계에서는 다양한 배차 환경에 대하여 10,000개의 에피소드를 실행하였으며, MAML을 활용하여 초기 정책을 최적화한 후 PPO를 적용하여 지속적인 정책 개선을 수행하였다. 시험 단계에서는 학습된 정책을 바탕으로 100개의 새로운 배차 시나리오를 설정하고, 각 시나리오에서의 성능을 평가하였다.

배차 성능 비교를 위해 다음의 4가지 평가 지표를 이용하였다.

- 시간 효율성(TE): 배차 완료까지 소요된 평균 시간
- 차량 활용도(VU): 배차된 차량의 평균 가동률
- 화물 처리량(CT): 단위 시간당 운송 완료된 화물 수
- 배차 균등성(DE): 차량별 배차 횟수의 분산

또한, 알고리즘의 전반적인 최적화 성능을 평가하기 위해 Ⅲ장에서 설명한 APM을 활용하였다.

5.2 배차 알고리즘 성능 비교

본 절에서는 모의실험을 통해 다양한 배차 환경에서 MAML+PPO 기반 알고리즘의 성능을 평가하고, 기존 GA, RBH 및 PPO 알고리즘과 비교·분석한다. 배차 시스템의 성능은 4.1절에서 설명한 4가지 핵심 지표를 통해 평가한다.

표 2는 각 배차 알고리즘별 주요 성능 지표 값을 나타내며, 이를 시각화하기 위해 그림 3의 막대그래프로 표현하였다. RBH와 GA에 비해 PPO 기반 알고리즘은

표 2. 알고리즘별 성능 비교 결과
Table 2. Performance comparison of dispatch algorithms.

Algorithm	RBH	GA	PPO	MAML +PPO
TE(%)	69.3	77.5	85.3	92.5
VU(%)	65.7	75.2	81.0	88.4
CT(%)	68.9	76.8	84.2	91.2
DE(%)	64.5	72.3	78.3	86.7

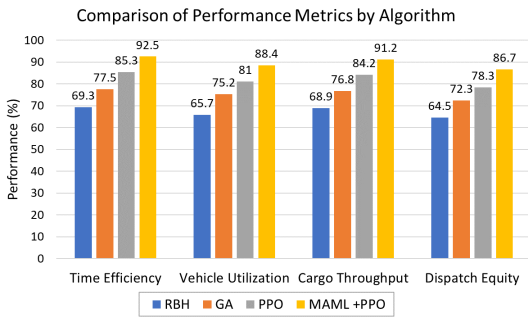


그림 3. 알고리즘별 성능 지표 비교 그래프
Fig. 3. Comparison of performance metrics by dispatch algorithm.

높은 시간 효율성(TE)을 보이며, 지속적인 정책 최적화를 통해 성능을 점진적으로 개선할 수 있다. 특히, MAML과 결합된 PPO 알고리즘은 초기 정책을 최적화함으로써 PPO 단독 모델보다 더 신속한 배차를 수행하며, 다양한 환경에서도 높은 적응력을 나타낸다.

반면, GA는 반복적인 탐색 과정을 거치면서 최적해를 찾아가므로 수렴 속도가 상대적으로 느린 단점이 있고, RBH는 정해진 규칙에 따라 배차가 이루어지므로 복잡한 환경 변화에 적절히 대응하지 못해 상대적으로 낮은 성능을 보였다.

MAML+PPO 알고리즘은 차량 활용도(VU) 측면에서도 GA 및 RBH보다 우수한 성능을 보였다. GA는 최적해를 찾는 과정에서 특정 차량에 배차가 몰리는 경향이 있으며, RBH는 사전 정의된 규칙을 따르다 보니 특정 차량이 반복적으로 선택되는 문제가 발생할 수 있다. 반면, MAML+PPO는 차량의 가용성을 고려하여 배차 균형을 유지함으로써 전체적인 차량 활용도를 향상시키는 결과를 얻을 수 있다.

화물 처리량(CT)과 배차 균등성(DE) 측면에서도 MAML+PPO 결합 알고리즘이 다른 기존 알고리즘보다 안정된 성능을 보이는 것을 확인할 수 있다. 제안한 알고리즘은 기존 알고리즘이 가지는 한계점을 극복하고, 환경 변화에 따라 배차 정책을 조정하여 균등하고 신속하게 배차 결정할 수 있는 기법임을 알 수 있다.

또한, 각 성능 지표가 종합 성능(APM)에 미치는 영향을 평가하기 위해 기본 시나리오를 기준으로 하여 성능 지표 가중치별 APM 성능을 표 3에 나타내었다. 그 결과, 시간 효율성(TE)과 차량 활용도(VU) 지표가 APM에 상대적으로 큰 영향을 미치는 것으로 나타났다. 이는 실시간 배차 성능 향상을 위해 시간 관리와 차량 활용 전략이 특히 중요함을 의미한다. 또한 배차 균등성(DE) 역시 화물 처리량(CT)보다 더 중요한 성능 지표

표 3. 성능 지표 가중치별 APM 비교
Table 3. The APM comparison of weighted performance metrics.

Scenarios	TE	VU	CT	DE	APM
Base scenario	0.25	0.25	0.25	0.25	0.86
TE emphasis	0.55	0.15	0.15	0.15	0.91
VU emphasis	0.15	0.55	0.15	0.15	0.89
CT emphasis	0.15	0.15	0.55	0.15	0.84
DE emphasis	0.15	0.15	0.15	0.55	0.87

임을 확인할 수 있다. 이는 제안한 알고리즘이 실시간으로 변화하는 운송 환경에 효과적으로 대응하여 배차의 효율성과 공정성을 높일 수 있음을 보여준다.

그림 4는 MAML+PPO, PPO, GA 및 RBH 알고리즘의 APM 값에 대한 누적 분포 함수(cumulative distribution function, CDF)를 나타낸다. RBH 알고리즘(검은색)은 가장 낮은 APM 값을 기록하였으며, 대부분의 경우 0.4 이하의 성능을 보였다. 이는 RBH가 사전에 정의된 규칙을 기반으로 배차를 수행하므로, 다양한 운송 환경 변화에 적응하기 어려운 한계를 가지기 때문이다. 또한, GA 알고리즘(파란색)은 RBH 대비 더 나은 성능을 보였지만, 여전히 0.6 이하에서 대부분의 APM 값이 분포하는 것을 확인할 수 있다. GA 기반 최적화 방식은 높은 연산 비용과 긴 계산 시간이 요구되므로, 실시간으로 변화하는 운송 환경에서는 활용하기 어려울 것으로 보인다. PPO 알고리즘(녹색)은 강화학습 기반으로 정책을 최적화함으로써 GA 및 RBH 대비 향상된 성능을 보였다. PPO는 정책 업데이트 과정에서 클리핑(clipping) 기법을 활용하여 안정적인 학습을 수행하고, 지속적인 학습을 통해 배차 정책을 개선할 수 있다. 이를 통해, GA 대비 더 높은 APM 값을 기록하며, 대부분의 배차 성능이 0.7 이상으로 분포하는 것을 확인할 수 있다. 마지막으로, MAML+PPO 알고리즘(빨간색)은 배차 알고리즘 중 가장 우수한 성능을 보였다. CDF 곡선이 가장 높은 APM 값을 기록하며, 대부분의 경우 0.85 이상의 성능을 달성하는 것으로 나타났다. 이는 MAML이 다양한 환경에서 빠르게 적응할 수 있도록 초기 정책을 최적화하고, PPO가 지속적인 정책 개선을 수행함으로써 배차 성능을 극대화할 수 있음을 알 수 있다. 특히, MAML을 활용하여 다양한 시나리오에서 사전 학습된 정책을 활용함으로써, PPO 단독 사용 대비 더욱 빠르게 최적의 배차 정책을 학습할 수 있다.

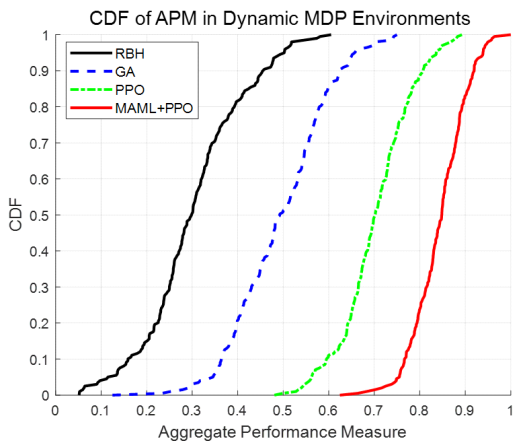


그림 4. CDF-APM을 이용한 배차 알고리즘 성능 비교
Fig. 4. Performance comparison of dispatch algorithms using CDF-APM.

본 논문에서 제안하는 MAML+PPO 알고리즘은 기존의 RBH 및 GA 기반 배차 알고리즘이 가진 한계를 극복하고, 환경 변화에 따라 배차 정책을 동적으로 조정할 수 있음을 실험을 통해 확인할 수 있다. 따라서 항만 물류 운송 시스템에서 다양한 배차 환경을 고려할 때, MAML과 PPO를 결합한 알고리즘을 적용하여 차량/화물을 효율적으로 배차할 수 있는 성능을 얻을 수 있다.

Ⅵ. 결 론

본 논문에서는 다양한 운송 환경에서 차량/화물 자동 배차의 최적화를 위해 MAML과 PPO를 결합한 강화학습 기반 배차 알고리즘을 제안하였다. 모의실험을 통해 MAML+PPO 기반 알고리즘의 배차 성능을 분석한 결과, 기존 GA 및 RBH 기반 최적화 기법 대비 시간 효율성, 차량 활용도, 화물 처리량, 배차 균등성에서 우수한 성능을 보였다. 특히, 제안된 알고리즘은 초기 정책 최적화를 통해 신속한 적응성을 확보하며, PPO 기반 정책 개선을 통해 지속적인 성능 향상이 가능함을 확인하였다. 이러한 결과에 따라, MAML과 PPO 결합 알고리즘 기반 배차 시스템이 실시간 운송 환경에서도 높은 최적화 성능을 제공할 수 있음을 검증하였다. 향후 연구에서는 다양한 메타 강화학습 기법을 추가 적용하여 배차 최적화 성능을 향상시키고, 실제 물류 데이터를 반영한 실시간 배차 환경에서 성능을 검증할 예정이다. 본 논문에서 제안하는 기법은 항만물류 시스템뿐만 아니라, 도시 물류 및 대규모 운송 네트워크에서의 자동 배차 최적화 기술 발전에 기여할 수 있을 것으로 기대된다.

References

[1] The Korea Transport Institute "*Global Logistics Technology Trend*," vol. 16, no. 675, Apr. 2022.

[2] S.-H. Kim, W. Lee, S.-H. Oh, J.-W. Lee, and W.-J. Kim, "The design of automation simulation system for efficient logistics management," *J. KIECS*, vol. 17, no. 1, pp. 187-192, Feb. 2022.

[3] M. Nazari, A. Oroojlooy, L. V. Snyder, and M. Takáč, "Reinforcement learning for solving the vehicle routing problem," in *Proc. 32nd Int. Conf. NeurIPS*, 2018.

[4] W. Kool, H. van Hoof, and M. Welling, "Attention, learn to solve routing problems," in *Proc. 7th ICLR*, 2019.

[5] H.-J. Lee, W.-S. Jang, S.-J. Lee, and D.-K. Kim, "Optimal dispatch method based on genetic algorithm in port logistics environment," in *Proc. KIIT Conf.*, pp. 236-239, Jeju, Korea, Nov. 2023.

[6] D.-H. Hong and G.-J. Kim, "Optimizing the vehicle dispatching for enhancing operation efficiency of container terminal," *J. Korea Convergence Soc.*, vol. 8, no. 10, pp. 19-28, Oct. 2017.

[7] T. Niine and O. Koppel, "*Competence in logistics - designing a meta-model of logistics knowledge areas*," Vienna, Austria: DAAAM International Scientific Book, 2014.

[8] J. J. Q. Yu, W. Yu, and J. Gu, "Online vehicle routing with neural combinatorial optimization and deep reinforcement learning," in *IEEE Trans. Intell. Transportation Syst.*, vol. 20, no. 10, pp. 3806-3817, Oct. 2019.

[9] Z. Zong, T. Feng, J. Wang, T. Xia, D. Jin, and Y. Li, "Deep reinforcement learning for demand driven services in logistics and transportation systems: A survey," *ACM Trans. Knowledge Discovery from Data*, vol. 19, no. 4, pp. 1-42, 2025.

[10] S. Ham, W. Hwang, and K. Lee, "Fast adaptation of network congestion prediction using meta learning," in *Proc. Winter Conf. KICS*, p. 613, 2020.

[11] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th ICML*, pp. 1126-1135, 2017.

[12] J.-H. Kim, D.-M. Shin, S.-E. Lee, and H.-N. Kim, "Performance analysis of vehicle/cargo transport matching algorithm based on MAML," *J. IEIE*, vol. 60, no. 9, Sep. 2023.

[13] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, Aug. 2017.

[14] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *Proc. 32nd Int. Conf. Mach. Learning, PMLR*, vol. 37, pp. 1889-1897, 2015.

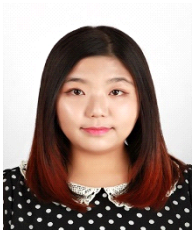
[15] Á. Belmonte-Baeza, et al., "Meta reinforcement learning for optimal design of legged robots," *IEEE Robotics and Automat. Lett.*, vol. 7, no. 4, pp. 12134-12141, 2022.

[16] X. Xu, X. Zeng, Z. Huang, and L. Xiao, "Meta reinforcement learning for adaptive resource allocation in wireless networks," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6202-6214, 2021.

[17] S. M. Jordan, et al., "Evaluating the performance of reinforcement learning algorithms," *Int. Conf. Machine Learn., PMLR*, pp. 4962-4973, 2020.

[18] Liangqu Long, "*MAML-Pytorch Imple-mentation*," GitHub repository, 2018, from https://github.com/dragen1860/MAML-pytorch

김 지 현 (Ji-Hyeon Kim)



2017년 8월 : 부산대학교 전자공학과 학사 졸업
2019년 2월 : 부산대학교 전기전자컴퓨터공학과 석사 졸업
2025년 8월 : 부산대학교 전기전자공학과 박사 졸업
2025년 9월~현재 : 부산대학교 컴퓨터 및 정보통신연구소 연수연구원
<관심분야> 적응신호처리, 레이더 신호처리, 배열 신호처리, 생체 신호처리
[ORCID:0000-0003-1425-2367]

권 순 영 (Soon-Young Kwon)

2018년 2월 : 부산대학교 전자공학과 학사 졸업
2018년 3월~현재 : 부산대학교 전기전자공학과 석박통합과정
<관심분야> 디지털 방송 신호처리, 레이더 신호처리
[ORCID:0000-0002-7280-8549]

신 다 민 (Da-Min Shin)

2024년 2월 : 부산대학교 전자공학과 학사 졸업
2024년 3월~현재 : 부산대학교 전기전자공학과 석사과정
<관심분야> 전자전·레이더 신호처리
[ORCID:0009-0004-0489-5072]

김 형 남 (Hyoung-Nam Kim)

1993년 2월 : 포항공과대학교 전자전기공학과 학사 졸업
1995년 2월 : 포항공과대학교 전자전기공학과 석사 졸업
2000년 2월 : 포항공과대학교 전자전기공학과 박사 졸업
2000년 5월 : 포항공과대학교 전자컴퓨터공학부 박사후연구원
2000년 5월~2003년 2월 : 한국전자통신연구원 무선방송연구소 선임연구원
2003년 3월~2007년 2월 : 부산대학교 전자전기통신공학부 조교수
2007년 3월~2012년 2월 : 부산대학교 전자전기통신공학부 부교수
2009년 2월~2010년 2월 : Johns Hopkins Univ. Visiting Scholar
2015년 9월~2016년 8월 : Univ. of Southampton Visiting Professor
2012년~현재 : 부산대학교 전자공학과 교수
<관심분야> 적응신호처리, 레이더 및 소나 신호처리, 디지털 방송신호처리, 생체신호처리
[ORCID:0000-0003-3841-448X]