

# 오프라인 데이터 기반 진화 알고리즘에 대한 소개

김은찬, 여진호, 최태종\*

전남대학교

kie8898@jnu.ac.kr, yjh4209@jnu.ac.kr, ctj17@jnu.ac.kr

## Introduction to Offline Data-driven Evolutionary Algorithms

Kim Eun Chan, Yeo Jin Ho, Tae Jong Choi

Chonnam National Univ.

### 요약

이 논문에서는 고비용 최적화 문제(Expensive Optimization Problem)를 해결하는데 최근 주목을 받고 있는 데이터 기반 진화 알고리즘(Data-driven Evolutionary Algorithms)에 대해서 소개한다. 데이터 기반 진화 알고리즘은 여러 가지 머신러닝 모델을 바탕으로 대리 모델(Surrogate Model)을 생성하여 최적화를 수행하는 방법이다. 특히, 데이터 기반 진화 최적화는 새로운 데이터의 확보 가능 여부에 따라 온라인 데이터 기반 진화 알고리즘과 오프라인 데이터 기반 진화 알고리즘으로 구분할 수 있다. 이 논문에서는 보다 범용적으로 활용될 수 있는 오프라인 데이터 기반 진화 알고리즘에 대해서 소개하고 그 한계와 극복 방법에 대해서 논의한다.

### I. 서론

최적화[1]란 주어진 목표를 달성하기 위해 최적해를 탐색하는 과정을 말한다. 최적화는 주로 공학, 경영, 경제, 과학 등 다양한 분야에서 중요한 역할을 맡고 있다. 또한 각 분야에 나타나는 문제를 해결하기 위해 다양한 기법들이 연구되어 오고 있다.

진화 알고리즘(Evolutionary Algorithm) [2]이란 자연의 진화 과정을 모티브로 한 최적화 기법이다. 이는 다양한 후보해(Candidate solution)를 생성하여 반복적으로 진화시키고, 적합도가 높은 후보해를 선택하여 변형함으로써 최적해에 탐색하는 기법이다. 진화 알고리즘은 내부 구조를 알 수 없는 블랙박스(Black-box) 형태의 복잡한 문제를 해결하는데 효과적인 방식이다. 진화 알고리즘의 경우 공학 설계, 게임 전략 개발, 초매개변수 탐색, 금융 투자 전략 등 다양하고 복잡한 문제를 해결할 수 있는 접근법으로 평가되고 있다 [2].

실세계 최적화 문제는 많은 시간과 높은 비용을 요구하는 고비용 최적화 문제(Expensive Optimization Problem)이다. 이러한 문제는 후보해를 평가하는데 막대한 양의 자원이 요구되기 때문에 기존의 진화 알고리즘으로 이러한 문제를 해결하는데 한계가 존재한다[3,4]. 따라서 많은 연구자들은 고비용 최적화 문제를 효과적으로 해결하기 위한 다양한 방법을 연구해왔다. 이 중에서 데이터 기반 진화 알고리즘(Data-driven Evolutionary Algorithms)은 최근 주목을 받고 있는 방법으로 기존에 축적된 데이터를 활용하여 대리 모델(Surrogate Model)을 구축하고 이를 기반으로 최적화를 수행한다[3,4]. 즉, 데이터 기반 진화 알고리즘은 후보해를 평가할 때 고비용 최적화 함수를 직접적으로 사용하는 것이 아니라 대리 모델을 통해 근사한 함수를 사용함으로써 후보해를 평가하는데 필요한 시간과 비용을 크게 절약할 수 있다. 오프라인 데이터 기반 진화 알고리즘은 주식 포트폴리오, 의료 의사 결정 지원, 항공기 설계 최적화 등에 성공적으로 적용되었다[4].

데이터 기반 진화 알고리즘은 기존에 수집된 데이터만을 활용하는 오프라인 데이터 기반 진화 알고리즘과 새로운 데이터를 수집하여 활용하는 온라인 데이터 기반 진화 알고리즘이 존재한다[4]. 온라인 데이터 기반 진화 알고리즘은 실시간으로 데이터를 활용하여 최적화를 수행하는 기법이다. 이는 변화하는 환경이나 실시간으로 반응해야 하는 상황에서 쓰이고 있다. 데이터 기반 진화 알고리즘은 금융 시장의 거래 전략 최적화, 실시간 제어 시스템 등에 활용되었다[4]. 특정 실세계 문제에서는 새로운 데이터를 수집하는 것이 불가능할 수 있다. 오프라인 데이터 기반 진화 알고리즘은 이러한 문제를 해결하기 위한 기법으로 사전에 수집된 데이터만을 활용하여 최적화 문제를 해결한다. 이 기법은 새로운 데이터를 사용할 수 없다는 점에서 최적화 문제에 가장 적합한 대리 모델을 사용하는 것은 매우 중요하다. 본 논문에서는 보다 범용적으로 활용될 수 있는 오프라인 데이터 기반 진화 알고리즘을 중심으로 소개하려고 한다.

표 1. 온라인 및 오프라인 데이터 기반 진화 알고리즘 비교

구분	온라인 데이터 기반 진화 최적화	오프라인 데이터 기반 진화 최적화
데이터 사용	실시간으로 수집되는 데이터 활용	고정된 데이터셋 활용
적용성	동적 환경 변화에 신속히 대응 가능	변화하는 환경에 대한 대응력 제한적
데이터 최신성	실시간 피드백으로 최신 정보 반영	과거 데이터에 의존, 최신성 저하
성능 개선	실시간 피드백을 통한 지속적 개선	데이터셋 고정으로 성능 개선 한계
적용 분야	실시간 의사결정이 필요한 문제	정적인 최적화 문제
복잡성	실시간 데이터 처리로 복잡도 증가	데이터 분석 과정이 상대적으로 단순

### II. 본론

#### 1. 오프라인 데이터 기반 진화 알고리즘 소개

오프라인 데이터 기반 진화 알고리즘이란 수집된 데이터만을 활용하여 최적화 문제를 해결하는 접근법이다. 오프라인 진화 최적화에서의 핵심은 고품질의 대리 모델을 구축해야 한다는 점이다. 여기서 대리 모델이란 실제로 우리가 해결하려는 복잡한 문제를 머신러닝 또는 딥러닝 등을 통해서 근사한 모델이다. 대리 모델을 구축하기 위해서는 데이터 수집 및 전처

리가 이루어져야 하며 모델 선택 및 구성이 이루어져야 한다. 주로 Polynomial response surface (PHS), Radial basis function (RBF), Support vector machines (SVM) 등이 대리 모델에 사용되고 있다[3]. 그 이후 수집된 데이터를 사용하여 각 대리 모델을 학습시키고, 교차 검증 기법을 통해 모델의 일반화 성능을 평가한 뒤 성능이 가장 좋은 모델을 선택한다. 마지막으로 최적화 수행을 하게 되면 대리 모델 구축하는 과정이 끝난다. 결과적으로 고품질의 대리 모델을 구축하는 것이 고비용 최적화 문제를 효과적으로 해결하는데 필수적이기 때문에 수집된 데이터를 기반으로 대리 모델을 효과적으로 학습시키고 적절한 모델을 선택해야 한다.

### 2. 오프라인 데이터 기반 진화 최적화 동작 과정

- ① **초기 데이터 샘플링:** 기존에 수집된 데이터에서 대리 모델을 학습하는데 활용할 데이터를 선택한다.
- ② **초기 대리 모델 구축:** 이 단계에서는 수집된 데이터 샘플들을 이용하여 대리 모델(surrogate model)을 구축하는 단계이다. 대리 모델을 선택할 때 다양한 종류의 대리 모델 중에서 적합한 모델을 선택한다.
- ③ **진화 알고리즘 기반 최적화:** 이 단계는 대리 모델을 사용하여 개체의 적합도를 평가하는 단계이다. 이는 진화 알고리즘을 통해 최적해를 탐색하며 고품질의 새로운 샘플 데이터를 선별한다.
- ④ **대리 모델 보강:** 이 단계에서는 선별된 새로운 샘플 데이터의 실제 목적 함수값을 계산하는 단계이다. 선별된 데이터를 이용하여 기존 대리 모델에 추가하여 대리 모델을 보강할 수 있다.
- ⑤ **반복 및 종료:** 이 단계에서는 3단계와 4단계를 반복 수행해야 하는 단계이다. 그리고 충분히 우수한 해를 발견하면 탐색을 종료한다.

### 3. 오프라인 데이터 기반 진화 알고리즘의 응용 사례

오프라인 데이터 기반 진화 알고리즘은 주식 포트폴리오 최적화에 적용되었다[5]. 이 연구에서 다루는 최적화 문제는 과거 주가 데이터를 활용하여 수익률과 위험을 최적화하기 때문에 실제 주가 데이터에 이상치, 결측치 등 품질에 대한 문제가 존재한다. 하지만 이 연구에서는 진처리를 활용하여 고품질 데이터를 확보하여 신뢰할 수 있는 최적의 포트폴리오를 도출하였다.

[6] 연구에서는 데이터 기반 진화 알고리즘을 활용하여 개인화된 의료 서비스를 제공하는 방법을 제안하였다. 이 연구에서는 오프라인 데이터 기반 진화 알고리즘을 활용하여 제한된 의료 데이터 안에서 데이터 마이닝 기술을 통해 핵심 정보와 주요 패턴을 추출하여 대리 모델을 생성하고 이를 활용하여 각 개인의 특성을 고려한 의사 결정을 지원하는 방법을 개발하였다.

[7] 연구에서는 오프라인 데이터 기반 진화 알고리즘을 항공기 설계 최적화에 적용하였다. 이 연구에서는 항공기 제조 기업이 보유하고 있는 데이터와 합성 데이터를 결합하여 대리 모델을 구축하였다.

## III. 결론

본 논문에서 오프라인 데이터 기반 진화 알고리즘을 중심으로 소개하였다. 오프라인 데이터 기반 진화 최적화는 과거에 수집된 데이터만을 활용하여 고비용 최적화 문제를 해결하는데 효과적인 접근법이다. 이 방법은 특히 데이터의 양이 제한적인 상황에서 매우 유용한 접근법이다. 오프라인 데이터 기반 진화 알고리즘은 주식 포트폴리오, 의료 의사 결정 지원, 항공기 설계 최적화 등에 성공적으로 적용되었다.

결과적으로, 오프라인 데이터 기반 진화 알고리즘은 고비용 최적화 문제와 관련된 다양한 산업과 연구 분야에서 중요한 역할을 맡고 있으며, 데이터 품질과 양의 문제를 보완하는 다양한 기술의 적용을 통해 더 신뢰할 수 있는 최적화 해결 방법을 제공할 수 있다.

향후 보다 대규모 최적화 및 다목적 최적화 등 복잡한 문제를 효과적으로 해결할 수 있는 방안을 연구할 계획이다. 또한 새롭게 개발된 기법들을 다양한 분야의 문제에 적용하고, 통합 활용방안을 모색할 계획이다.

## ACKNOWLEDGMENT

이 결과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2023-00214326, RS-2023-00242528).

## 참 고 문 헌

- [1] Chong, Edwin KP, and Stanislaw H. Żak. An introduction to optimization. Vol. 75. John Wiley & Sons, 2013.
- [2] Eiben, Agoston E., and James E. Smith. Introduction to evolutionary computing. Springer-Verlag Berlin Heidelberg, 2015.
- [3] He, Chunlin, et al. "A review of surrogate-assisted evolutionary algorithms for expensive optimization problems." *Expert Systems with Applications* 217 (2023): 119495.
- [4] Jin, Yaochu, et al. "Data-driven evolutionary optimization: An overview and case studies." *IEEE Transactions on Evolutionary Computation* 23.3 (2018): 442-458.
- [5] He, Cheng, et al. "A repository of real-world datasets for data-driven evolutionary multiobjective optimization." *Complex & Intelligent Systems* 6.1 (2020): 189-197.
- [6] Wang, Handing, Yaochu Jin, and Jan O. Jansen. "Data-driven surrogate-assisted multiobjective evolutionary optimization of a trauma system." *IEEE Transactions on Evolutionary Computation* 20.6 (2016): 939-952.
- [7] Li, Jian-Yu, et al. "Data-driven evolutionary algorithm with perturbation-based ensemble surrogates." *IEEE Transactions on Cybernetics* 51.8 (2020): 3925-3937.