강화 학습 기반의 Heuristic Evolution 시간 단축 방법론

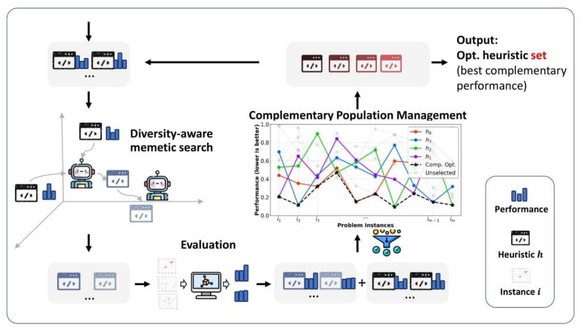
장현종[1], 김영훈[2]
[1]경희대학교 인공지능학과, [2]경희대학교 산업경영공학과

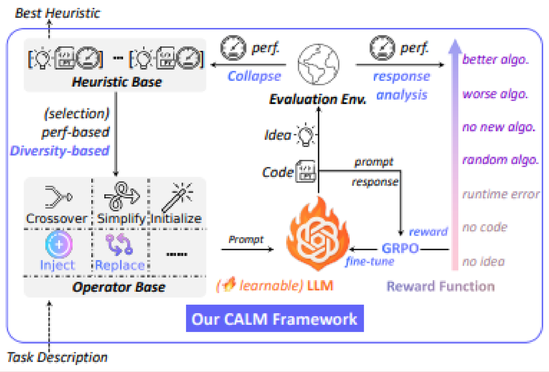
1. Introduction

- **Heuristic evolution 시간 단축 필요성**
  - ❖ 최근 대규모 언어 모델을 활용한 진화 휴리스틱(EoH) 연구는 단일 목적 최적화에 주로 집중되어 있음.
  - ❖ 실데이터가 있어도 **설계(프롬프트) 단계에서 직접 참조하지 못해** 데이터의 **분포·맥락을 반영하지 못함**.
  - ❖ 데이터셋 일부(다변량 변수)와 빈 용량 정보를 활용해 데이터셋의 분포·맥락을 설계에 반영하려면, 배치별 조건 변화로 인해 **EoH를 실시간으로 실행**해야 하는 제약이 발생함.
  - ❖ 이를 해결하기 위해 **강화학습으로 유망 연산자만 선택**해 수식 진화 과정을 효율화하고, **총 진화 시간**을 단축하는 접근을 제안함.

2. Related Works

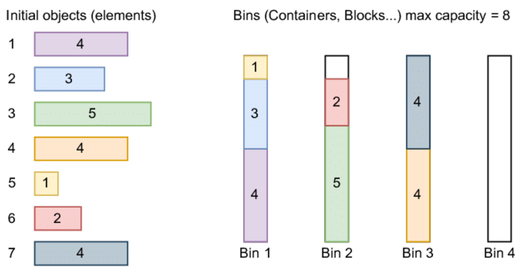
- **Evolution of Heuristics(EoH) 기반 자동 알고리즘 설계 선행 연구**
  - ❖ **FunSearch: LLM–평가기 공동 루프 기반의 휴리스틱 진화 프레임워크**
    LLM과 평가기를 반복 루프로 연결해 휴리스틱을 진화시키는 프레임워크. 평가기가 LLM의 환각을 억제하며 안정적으로 성능 향상을 이끌고, 온라인 빈 패킹에서도 FF/BF를 상회하는 결과를 달성함.
  - ❖ **EoH (Evolution of Heuristics): 대규모 언어모델 기반 휴리스틱 자동 진화**
    LLM이 생성한 사고(Thought)와 실행 코드(Code)를 진화 연산으로 개선하는 방식. 기존 FunSearch나 수작업 휴리스틱보다 계산 효율성과 성능 모두 우수함.
  - ❖ **EoH-S (Evolution of Heuristic Set): 다중 휴리스틱 협업 최적화**
    단일 휴리스틱의 한계를 극복하기 위해 상호보완적 휴리스틱 집합을 자동 설계. 다양한 인스턴스에서 강건성과 최대 60% 수준 성능 개선을 보여줌.
  - ❖ **CALM (Co-evolution of Algorithms and Language Model):알고리즘–언어모델 동진화(Co-evolution) 기반 성능 주도형 LLM 최적화**
    CALM: 휴리스틱 성능 신호를 활용해 LLM 자체를 강화학습으로 동진화(Co-evolution)하는 접근. 탐색 과정과 모델을 동시에 최적화하여 안정적인 성능을 확보함.





3. Background

- **Online Bin-Packing Problem**
  - ❖ 아이템이 순차 도착하며, 목표는 **사용 빈(bins) 수 최소화**.
  - ❖ 각 아이템은 크기를 가지고 있으며, 빈의 용량을 초과하지 않도록 빈에 배치해야 함.
  - ❖ 이 문제는 **NP-hard 문제**로 분류되며, 최적해를 찾는 것이 매우 어려움.
  - ❖ 목표는 가능한 적은 수의 빈을 사용하여 모든 아이템을 효율적으로 배치하는 것



연산자	설명
i1	백지에서 시작해 완전히 새로운 점수 함수를 설계·구현
e1	여러 부모 알고리즘을 보고, 형태가 다른 전혀 새로운 알고리즘 설계
e2	부모들의 공통 핵심 아이디어를 뽑아 그 뼈대로 새 알고리즘 설계
m1	한 부모를 기반으로 형태만 크게 바꿔 재 작성
m2	한 부모의 핵심 파라미터를 찾아 값/설정만 바꿔 변형
m3	과적합 위험·복잡한 부분을 덜어내 간단하게 재 작성
c1	기존 알고리즘에 새 컴포넌트 하나를 삽입해 기능 확장
c2	기존의 규칙/상수/모듈을 더 나은 것으로 교체
c3	여러 부모의 장점을 융합해 하나의 알고리즘으로 재 작성
c4	불필요한 요소를 제거해 더 단순·가벼운 형태로 재 작성

4. Methodology

- **배치 프로파일 주입 :** 설계(프롬프트) 단계에서 배치의 분포·맥락을 LLM이 참고하도록 요청하여, 초기 설계가 데이터에 **정합되도록 유도**하고 **성능 향상**을 꾀함.

❖ $$s_{\text{batch}} = \Big[\underbrace{C}_{\text{Capacity(default 100)}}, \underbrace{N}_{\text{Items}}, \underbrace{\mu}_{\text{Mean}}, \underbrace{\sigma}_{\text{Std.dev.}}, \underbrace{x_{\min}}_{\text{Min}}, \underbrace{x_{\max}}_{\text{Max}}, \underbrace{Q_{0.50}}_{\text{Median(p50)}}, \underbrace{Q_{0.90}}_{\text{Quantile(p90)}}, \underbrace{Q_{0.99}}_{\text{Quantile(p99)}}, \underbrace{r_{0.80}}_{\text{Tail ratio}(\geq 80\%)}, \underbrace{r_{0.90}}_{\text{Tail ratio}(\geq 90\%)}, \underbrace{(v_{(1)}, \ldots, v_{(k)})}_{\text{Top-}k\text{ frequent sizes}} \Big]$$

- **RL 최적 연산자 선택 :** 매 세대에 **유망 연산자만 선택**해 불필요한 변이/교차를 줄이고 **진화 시간을 단축**.

❖ $$s_t = \big[\underbrace{\text{best}_t}_{\text{Best P(Excess)}}, \underbrace{\text{avg}_t}_{\text{Mean P}}, \underbrace{\text{diversity}_t}_{\text{Population diversity}}, \underbrace{\text{popSize}_t}_{\text{Population size}}, \underbrace{\text{genTime}_t}_{\text{Generation runtime}} \big]$$

❖ $$\mathcal{A}_t = \begin{cases} \underbrace{\text{TopK}_k(\{Q(s_t, a) : a \in \mathcal{O}\})}_{\text{Top-}k\text{operators(exploitation)}}, & \text{with prob. } \underbrace{1-\varepsilon}_{\text{Exploitation prob.}} \\ \underbrace{\text{Uniform}_k(\mathcal{O})}_{\text{Random}k\text{operators(exploration)}}, & \text{with prob. } \underbrace{\varepsilon}_{\text{Exploration prob.}} \end{cases}$$

❖ $$r_t = \underbrace{\alpha\,(\text{best}_{t-1} - \text{best}_t)}_{\text{Performance improvement(Excess decrease)}} - \underbrace{\beta\,\text{Time}_t}_{\text{Generation-time penalty}}$$

❖ $$L_t(\theta) = \mathbb{E}\left[\left(\underbrace{r_t + \delta \max_{a'} Q_{\theta^-}(s_{t+1}, a')}_{\text{TD target}} - \underbrace{Q_\theta(s_t, a_t)}_{\text{Current estimate}} \right)^2 \right]$$

- 행동 선택 : 현재 s에서 상위 연산자 3개를 선택하거나 랜덤으로 선택
- 학습 : Q예측이 실제 보상+Target Network에 가깝도록 업데이트
- 스케줄 : 매 세대마다 새 전이를 버퍼에 추가하고 업데이트
- **Excess와 수행 시간을 DQN+ε-greedy 정책을 활용해 낮추는 것이 목적**

- **하이퍼파라미터 세팅**

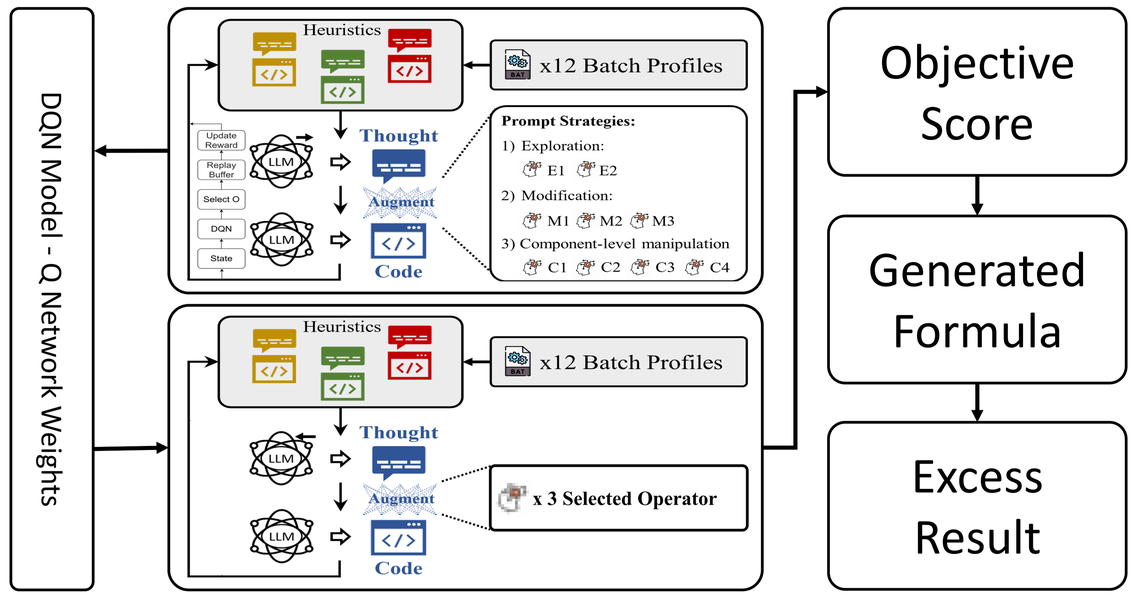
❖ **배치 프로파일 하이퍼파라미터 값**

Item	Value
N_scenarios	24
Top_k_bins	4
Capacity	{100, 300, 500}
Size	{1k, 5k, 10k}

❖ **DQN 강화학습 하이퍼파라미터 값**

Item	Value
LLM Model	openai api, `gpt-3.5-turbo-1106`
Learning Rate	$lr = 0.001$
Discount Factor	$\gamma = 0.99$
Initial Epsilon	$\epsilon = 0.2$
Epsilon Decay	$\epsilon \leftarrow 0.95 \times \epsilon$ (end of each run)
Batch Size	`batch_size = 32`
Operators Used	`e1, m1, c4`

- **강화 학습 기반 heuristic evolution 시간 단축 방법론 파이프라인**



5. Result

- **배치 프로파일 주입 30회 실험 평균 결과 비교**

❖ **12개 배치 프로파일을 참조하도록 LLM에 요청**

- 기존 방식(EoH)과 비교해, 신규 방식(batch+EoH)은 Capacity(빈의 용량)가 배치와 유사할 경우 높은 성능을 보이고, 유사하지 않을 경우 비교적 성능이 높지 않음

$$[\text{Excess} = \frac{\text{Number of bins used} \times \text{Bin capacity} - \text{Total item size}}{\text{Total item size}} \times 100\%]$$

<기존 EoH 실행 시 Excess(%)>

Dataset	Capacity	Excess (%)
Weibull 1k	100	3.78
Weibull 5k	100	2.78
Weibull 10k	100	2.63
Weibull 1k	300	3.44
Weibull 5k	300	2.63
Weibull 10k	300	2.48
Weibull 1k	500	3.49
Weibull 5k	500	2.69
Weibull 10k	500	2.54

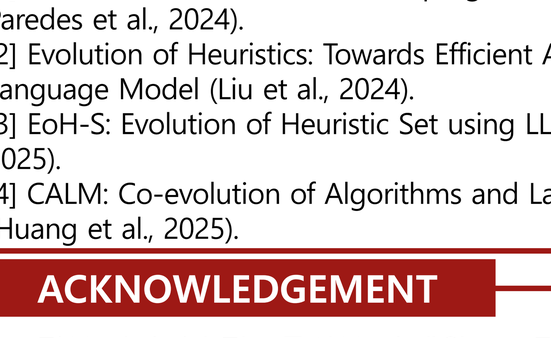
<배치 주입하며 EoH 실행 시 Excess(%)>

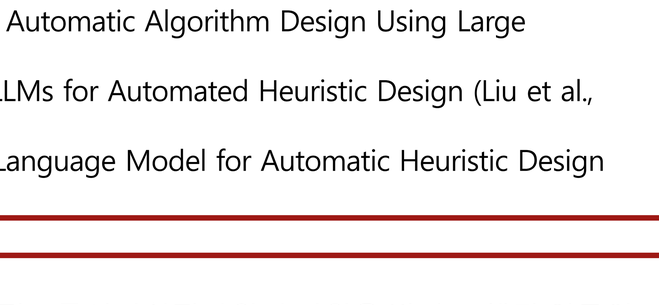
Dataset	Capacity	Excess (%)
Weibull 1k	100	2.12
Weibull 5k	100	0.84
Weibull 10k	100	0.51
Weibull 1k	300	0.78
Weibull 5k	300	0.69
Weibull 10k	300	0.58
Weibull 1k	500	4.31
Weibull 5k	500	6.29
Weibull 10k	500	7.31

- **RL 최적 연산자 선택 30회 실험 평균 결과**

❖ **DQN 강화학습 알고리즘 적용**

- 기존 방식(bponline_5o)과 비교해, 신규 방식(bponline_3o_RL)은 성능 차이를 0.01 이하로 유지하면서 진화 시간을 약 37.16% 단축함





CONCLUSIONS

1. Data-consistent design: 이 솔루션은 배치 프로파일 주입(통계·결정 예시 참조)과 강화학습 기반 최적 연산자 선택을 결합해, 설계 과정에서 데이터 정합성을 높이고 온라인 빈 패킹 문제에서 실용적인 시간-성능 균형을 달성함.
2. Batch Profile Injection : 30회 평균 실험에서 배치 프로파일 주입은 참조 프로파일이 대상 배치의 분포/용량과 유사할수록 Excess가 유의미하게 감소하는 것을 확인함.
3. Reduced evolution time : RL-OpSel(DQN) 은 기존 EoH 대비 약 37.16% 감소에도 불구하고 성능 차이를 ≤0.01로 유지하는 것을 확인함.

REFERENCE

[1] Mathematical discoveries from program search with large language models (Romera-Paredes et al., 2024).
[2] Evolution of Heuristics: Towards Efficient Automatic Algorithm Design Using Large Language Model (Liu et al., 2024).
[3] EoH-S: Evolution of Heuristic Set using LLMs for Automated Heuristic Design (Liu et al., 2025).
[4] CALM: Co-evolution of Algorithms and Language Model for Automatic Heuristic Design (Huang et al., 2025).

ACKNOWLEDGEMENT

정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임
(No.RS-2022-00155911, 인공지능융합혁신인재양성(경희대학교)).