ISCAS 中国科学院软件研究所学术年会
暨重点实验室科技活动周
2025 第十届
学术论文

基于约束引导对抗型智能体训练实现博弈对抗智能体的多样性测试

马序言[1,2]，王亚文[1,2]，王俊杰[1,2]，谢肖飞[3]，吴泊逾[1,2]，闫熠光[1,2]，李守斌[1,2]，徐帆江[1,2]，王青[1,2]

[1]中国科学院软件研究所 [2]中国科学院大学 [3]新加坡管理大学

IEEE Transactions on Software Engineering (TSE), 51(1), 66-81, 2025

联系方式：马序言，王亚文，王俊杰，王青 {maxuyan2021, yawen2018, junjie, wq}@iscas.ac.cn

背景和动机

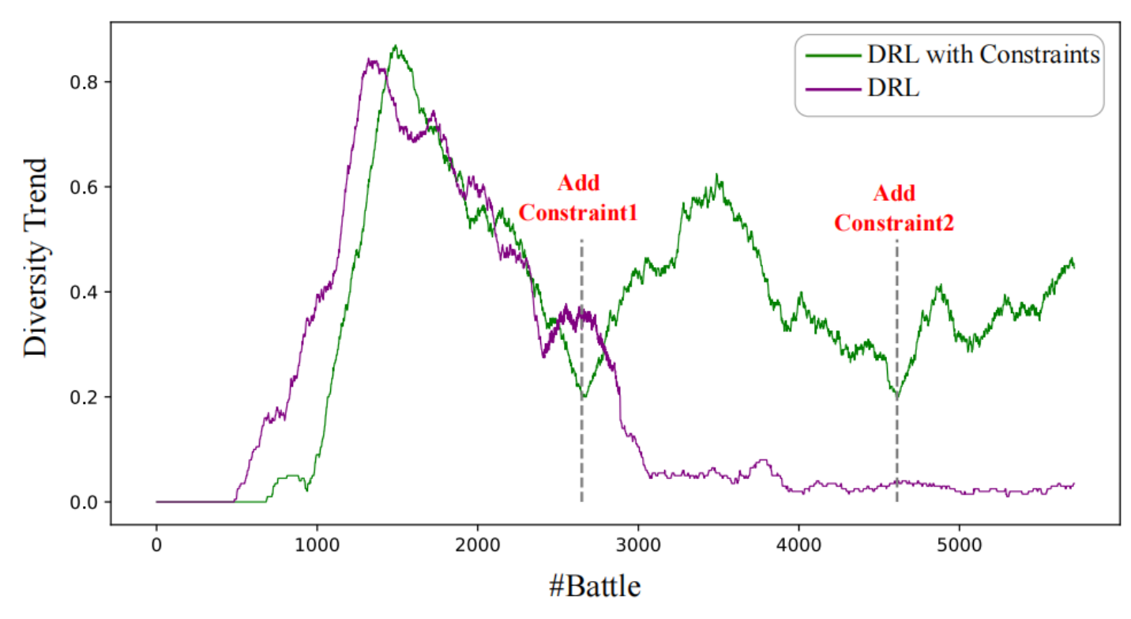
- ➢ 智能体的训练过程错综复杂，智能体不仅要适应静态规则，还要应对对手不断变化的对手、场景等。
- ➢ 在军事冲突或商业竞争等安全至关重要的环境中，不具备鲁棒性的代理可能会造成严重后果，从财务损失到人员生命安全风险。
- ➢ 多智能体系统运行过程中会产生大量数据，蕴含丰富的知识

现有方法的不足

◆已有基于变异、搜索等测试技术，在探索完部分状态空间后，缺乏有效指引，很难再探索到新状态

解决思路

◆通过数据智能分析，抽取显式约束，在博弈对抗过程中添加显式约束，引导智能体多样性的行为模式



通过限制对抗双方血量差不能过大，可以引导产生平缓的战胜模式

贡献

- 据我们所知，这是首次尝试利用约束训练对抗智能体的概念作为博弈对抗测试智能体的明确指导。
- 提出一个面向多样性的测试框架 AdvTest，采用约束引导对抗智能体训练，解决了以下挑战：哪些是合适的约束、何时引入约束以及需要添加哪些约束。
- 在《星际争霸 II》的四张地图上对 AdvTest 的有效性进行了实验评估，其性能表现优异，优于其他基线方法。

方法

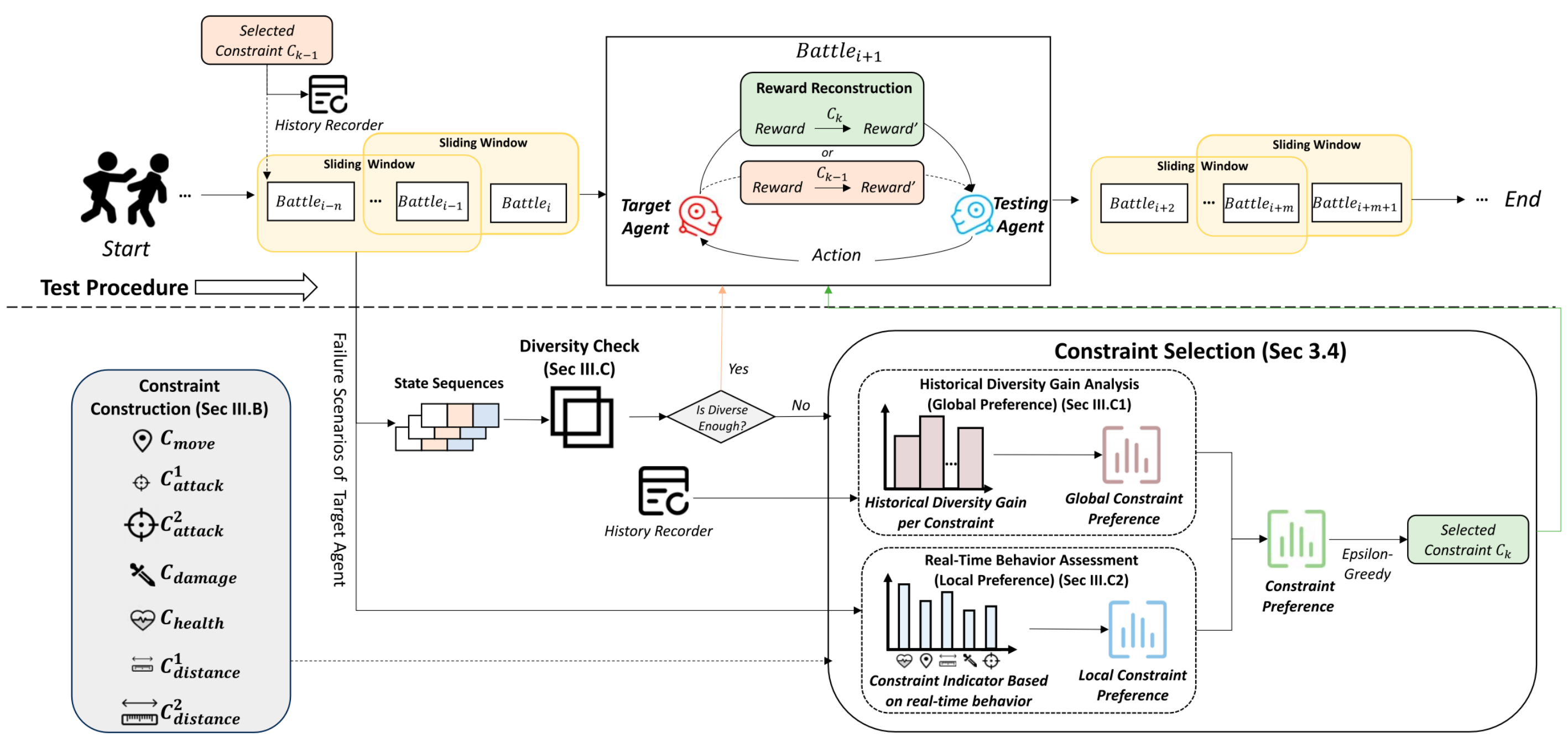
● 约束集

五个因素

智能体自身能力
  移动范围
  攻击范围
  伤害值
智能体交互状态
  生命值差异
  距离

七个约束

较小范围内移动C_{move}
较小范围内攻击C^1_{attack}
较大范围内攻击C^2_{attack}
较小攻击能力C_{damage}
生命值差异较小C_{health}
智能体距离较小C^1_{dist}
智能体具体较大C^2_{dist}



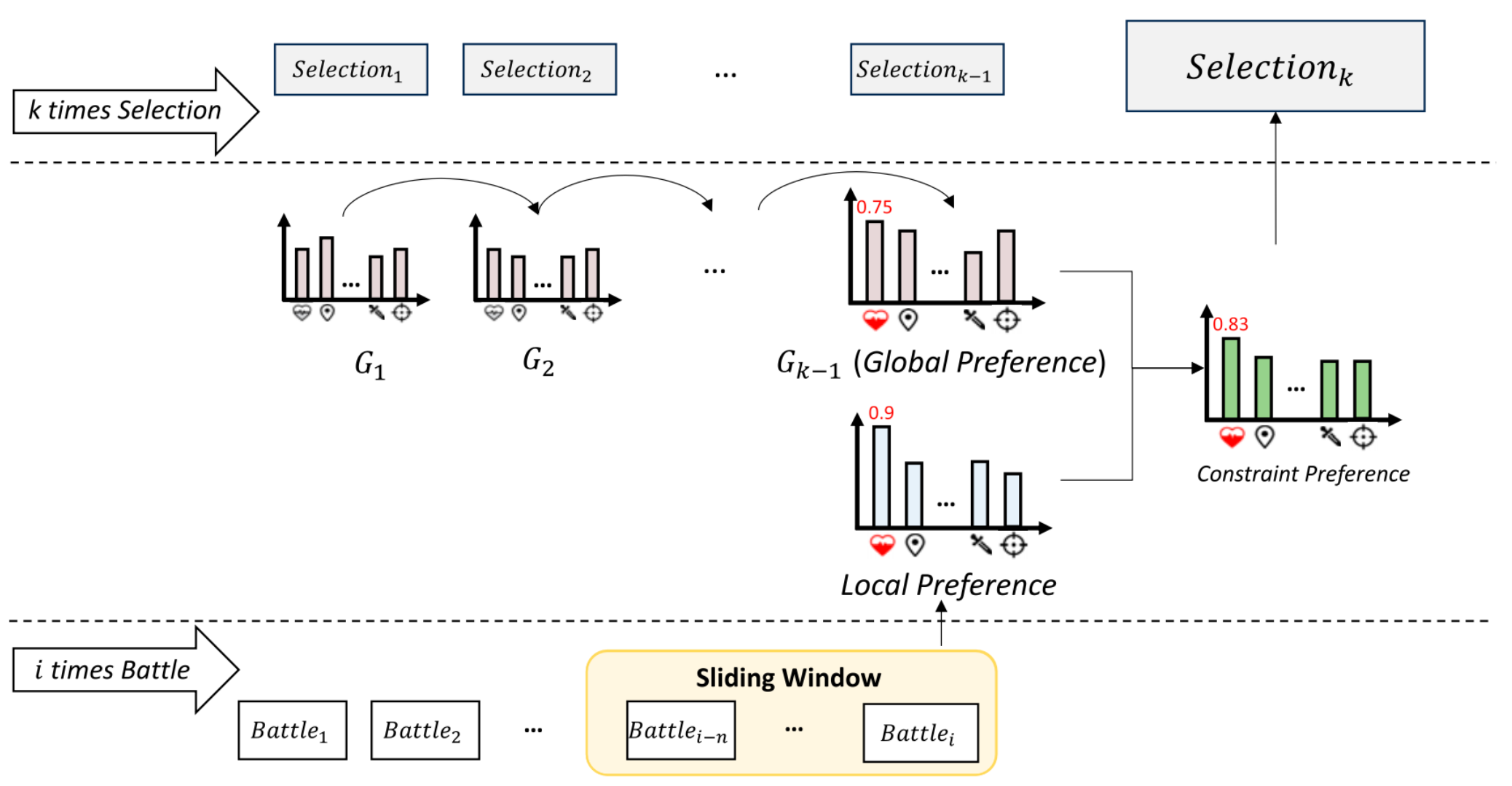
● 多样性检查

- ➢ 本模块的目标是衡量最近的失败场景与测试开始以来的现有场景是否不同。如果最近的失败场景多样性不足，这意味着测试代理以类似的方式击败目标代理，则 AdvTest 将转向“约束选择”以添加另一个约束来增强多样性。
- ➢ 设计了一个滑动窗口，来捕捉最近一段时间内的对战。状态序列可以表示为$Q = \{s_1, s_2, \ldots, s_w\}$，$w$为滑动窗口的大小。计算两个序列的距离的公式如下：

$$\mathrm{dis}(s, s') = \frac{Hamming(s, s') + |len(s) - len(s')|}{\max(len(s), len(s'))}$$

● 约束选择

- ➢ 全局偏好：收集全局信息，其中包含每次约束选择的历史记录以及每次选择带来的相应多样性增益。AdvTest 的目标是通过一定的约束选择策略来最大化累积多样性。将此任务类比为多臂老虎机问题，该问题最初的目标是通过特定的拉臂策略来最大化累积奖励。
- ➢ 局部偏好：记录实时约束违规情况，以计算最近的约束指标。让 AdvTest 根据相应的指标选择最近严重违反的约束。

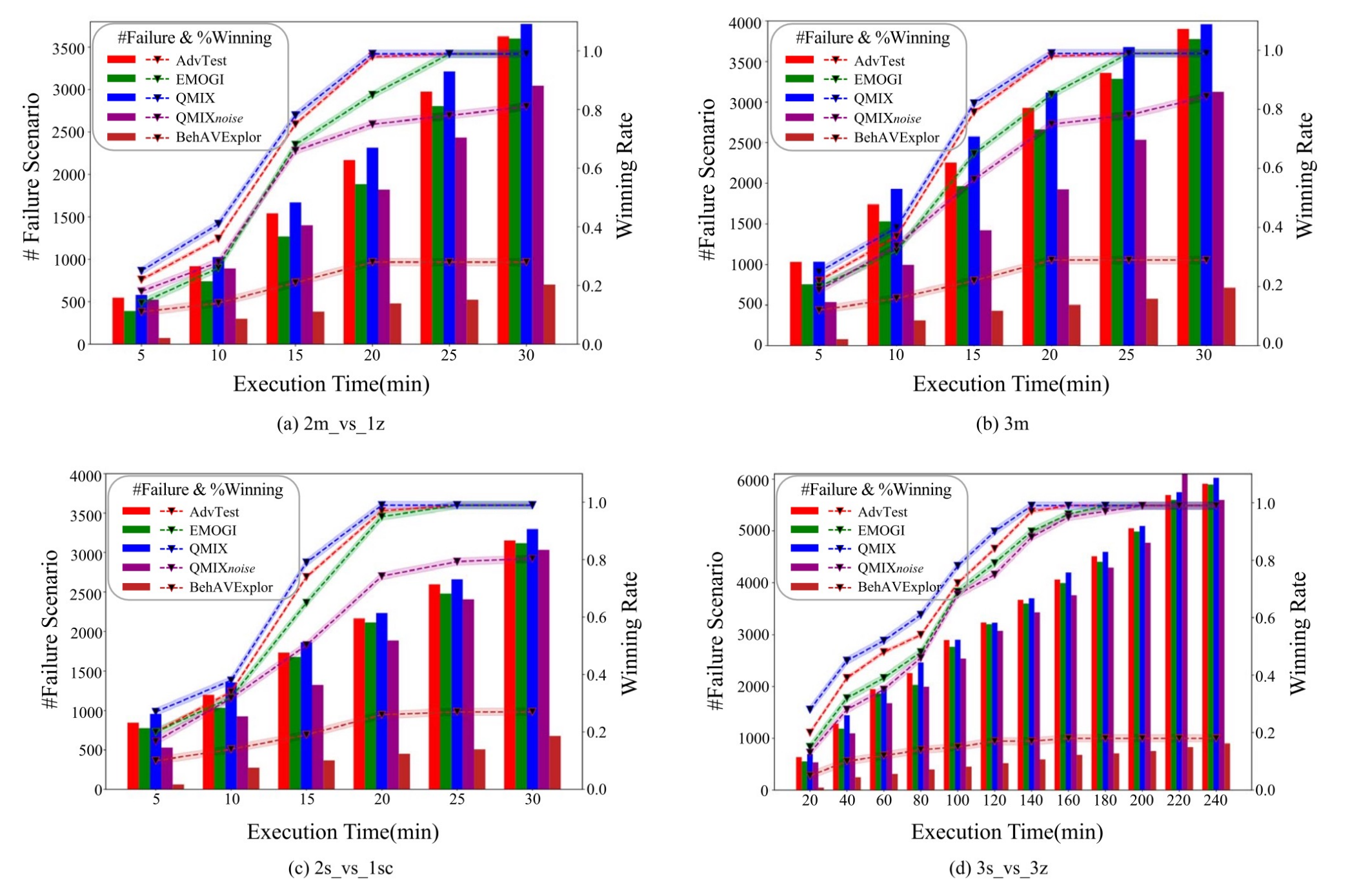


Algorithm 1 Constraint Selection with Global Preference

```
Input: Constraint Set $C$, the $T^{th}$ Selection
Output: Rewards $R(c_i)$
1: $\forall c_i \in \{c_1, \ldots, c_n\}$, $R(c_i) = 0$, $count(c_i) = 0$
2: for t = 1, 2,..., T do
3:     $c = \epsilon$-GreedySelection()
4:     $v = Reward(c)$
5:     $R(c) = \frac{R(c)*count(c)+v}{count(c)+1}$
6:     $count(c) = count(c) + 1$
7: end for
8: return $R(c_i)$
```

实验

➢ 胜率和失效场景数量的结果

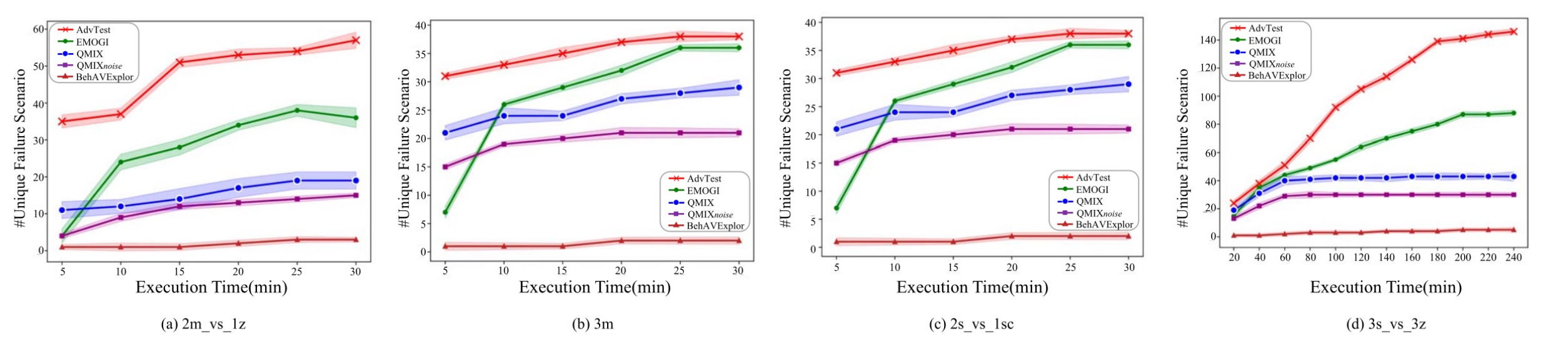


(a) 2m_vs_1z (b) 3m (c) 2s_vs_1sc (d) 3s_vs_3z

➢ 平均距离得分和状态覆盖率的结果

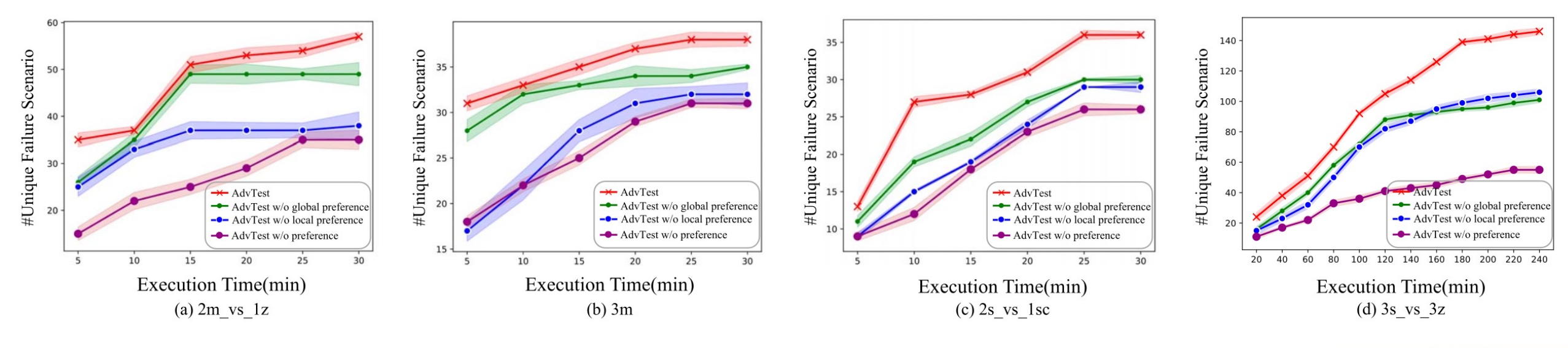
Map Name Method \ Metric	2m_vs_1z #Distance	2m_vs_1z %Coverage	3m #Distance	3m %Coverage	2s_vs_1sc #Distance	2s_vs_1sc %Coverage	3s_vs_3z #Distance	3s_vs_3z %Coverage
QMIX	0.0656	34.67%	0.1720	25.49%	0.0348	46.22%	0.4096	21.07%
QMIX$_{noise}$	0.0694	32.33%	0.1764	25.62%	0.0329	42.76%	0.3816	18.66%
EMOGI	0.1022	50.08%	0.2894	42.17%	0.0711	62.20%	0.5064	39.66%
BehAVExplor	0.0431	4.08%	0.1206	2.69%	0.0279	6.24%	0.3428	1.93%
AdvTest	**0.1924**	**77.21%**	**0.3341**	**61.49%**	**0.1235**	**82.17%**	**0.5929**	**56.41%**

➢ 唯一失效场景数量的结果



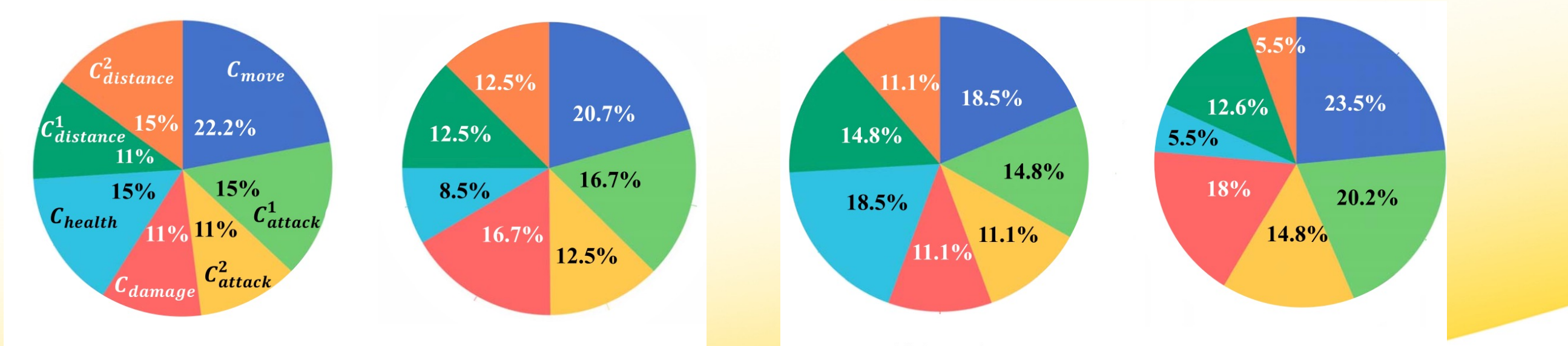
(a) 2m_vs_1z (b) 3m (c) 2s_vs_1sc (d) 3s_vs_3z

➢ 消融实验结果



(a) 2m_vs_1z (b) 3m (c) 2s_vs_1sc (d) 3s_vs_3z

➢ 约束的必要性评估



(a) 2m_vs_1z (b) 3m (c) 2s_vs_1sc (d) 3s_vs_3z