北京大学
PEKING UNIVERSITY

群体机器人研究前沿

——研究生在线学术会议

谭营 教授

北京大学 信息科学技术学院 智能科学系

ytan@pku.edu.cn

2020.11.07

北京大学 计算智能实验室
Computational Intelligence Laboratory, Peking University

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- 基于三角编队搜索的多目标搜索策略
- 基于概率有限状态机的多目标搜索策略
- 基于深度学习和进化计算的多目标搜索策略
- 群体协作在游戏AI中的应用
- 总结与展望

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- 基于三角编队搜索的多目标搜索策略
- 基于概率有限状态机的多目标搜索策略
- 基于深度学习和进化计算的多目标搜索策略
- 群体协作在游戏AI中的应用
- 总结与展望

群体协同

生物群体中的协同行为

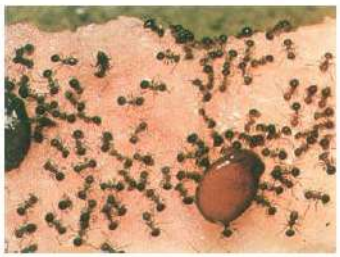
个体简单　局部信息交互　群体规模变化大　群体行为复杂

群体智能

针对生物群体的仿生　强调简单个体间协同

群体机器人

群体智能+机器人



群体机器人的定义

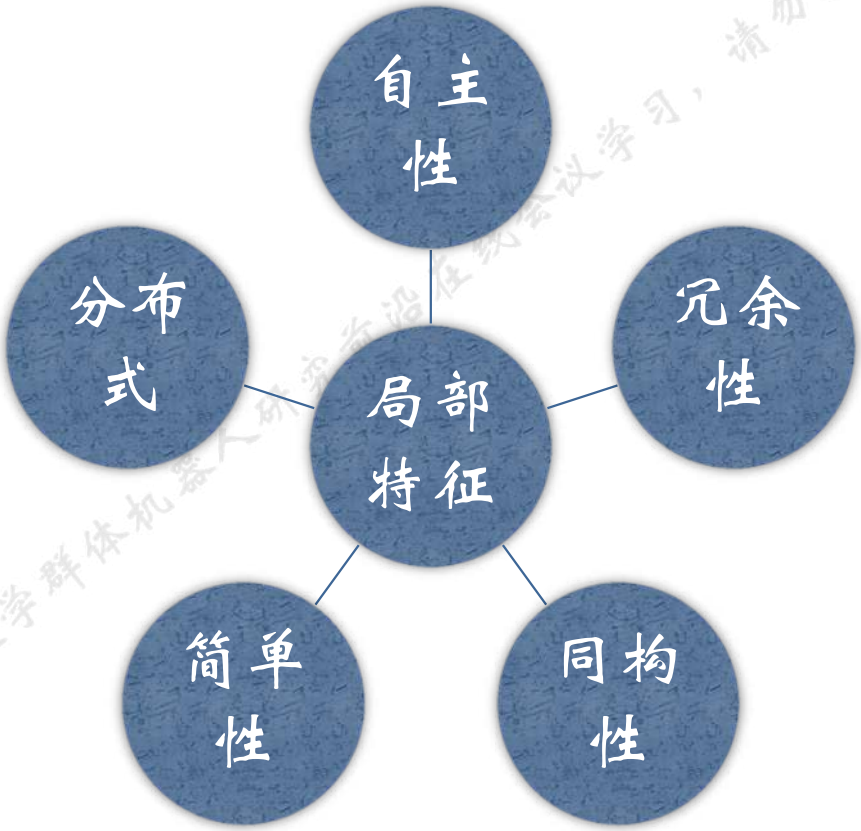
- 启发自社会性生物群体：
  - 鸟类迁徙、蚂蚁觅食等
  - 个体较弱
  - 群体行为复杂

- 群体机器人
  - 大量、简单的实物智能体
  - 集体行为从局部交互涌现

群体机器人的局部特征

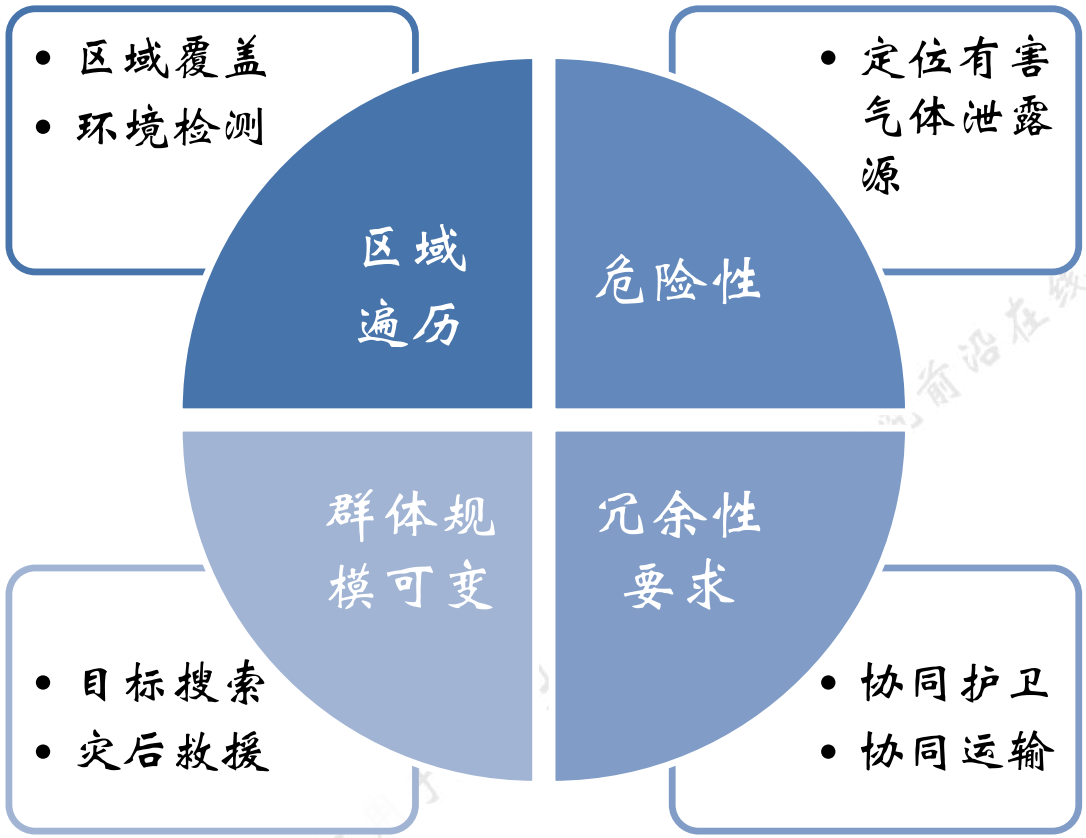
交互

控制



群体机器人的系统属性

群体机器人的应用领域



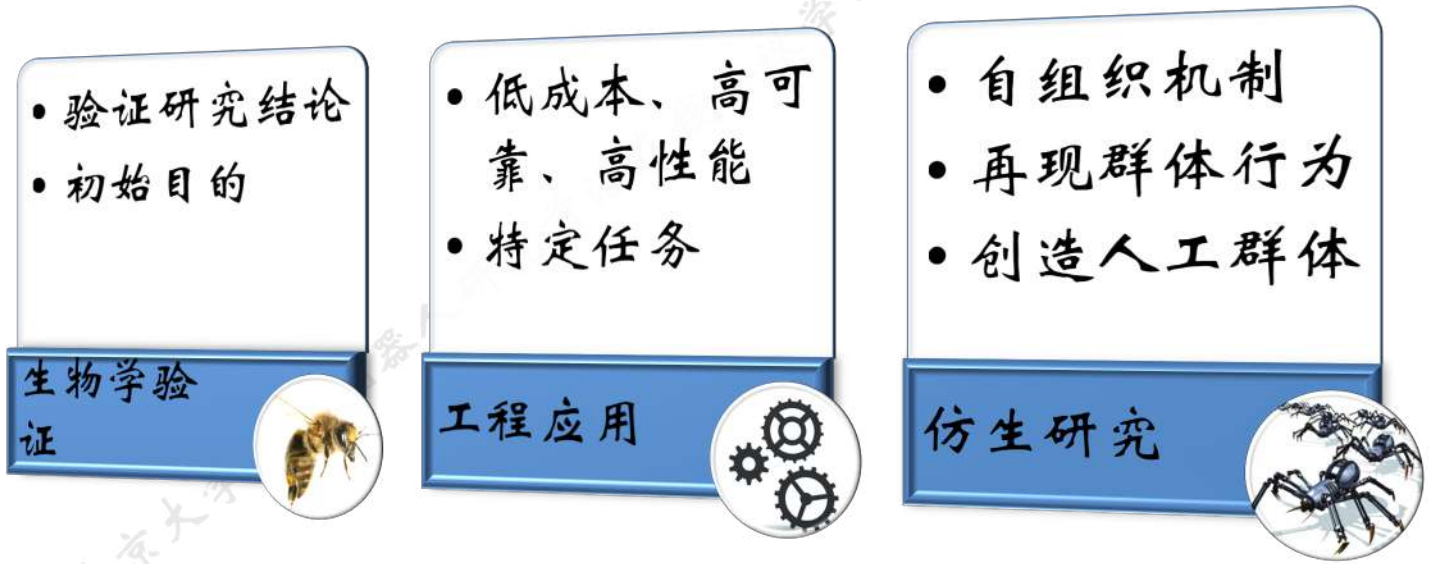
需要大量个体

- 区域覆盖
  - 区域巡逻、搜救
  - 水质监测
- 群体组织
  - 图形生成
  - 区域封锁

危险性任务

- 区域覆盖
  - 排雷、矿道清理
  - 废墟救援
- 群体组织
  - 区域防御

群体机器人的研究意义

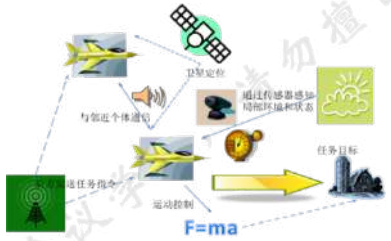


大纲

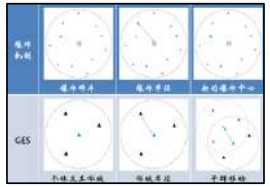
- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- 基于三角编队搜索的多目标搜索策略
- 基于概率有限状态机的多目标搜索策略
- 基于深度学习和进化计算的多目标搜索策略
- 群体协作在游戏AI中的应用
- 总结与展望

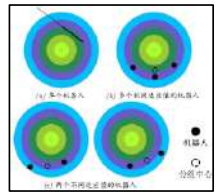
群体机器人多目标搜索问题

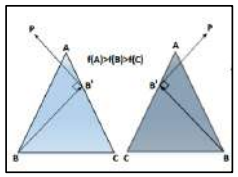
- 群体机器人多目标搜索
  - 一群机器人在广阔未知的环境中，通过某种协同机制，搜索与处理环境中的目标



抽象建模



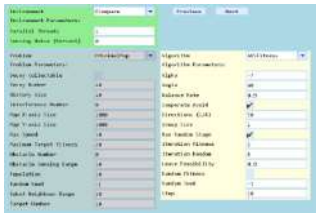


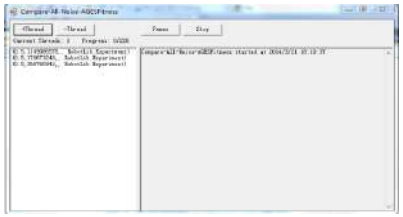


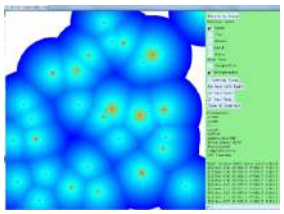
提出多个多目标协同搜索方法

- 衡量指标
  - 搜索与处理所有目标所需时间

- 应用前景
  - 民用：灾后救援、水质监测
  - 军用：敌潜搜寻、战场监测



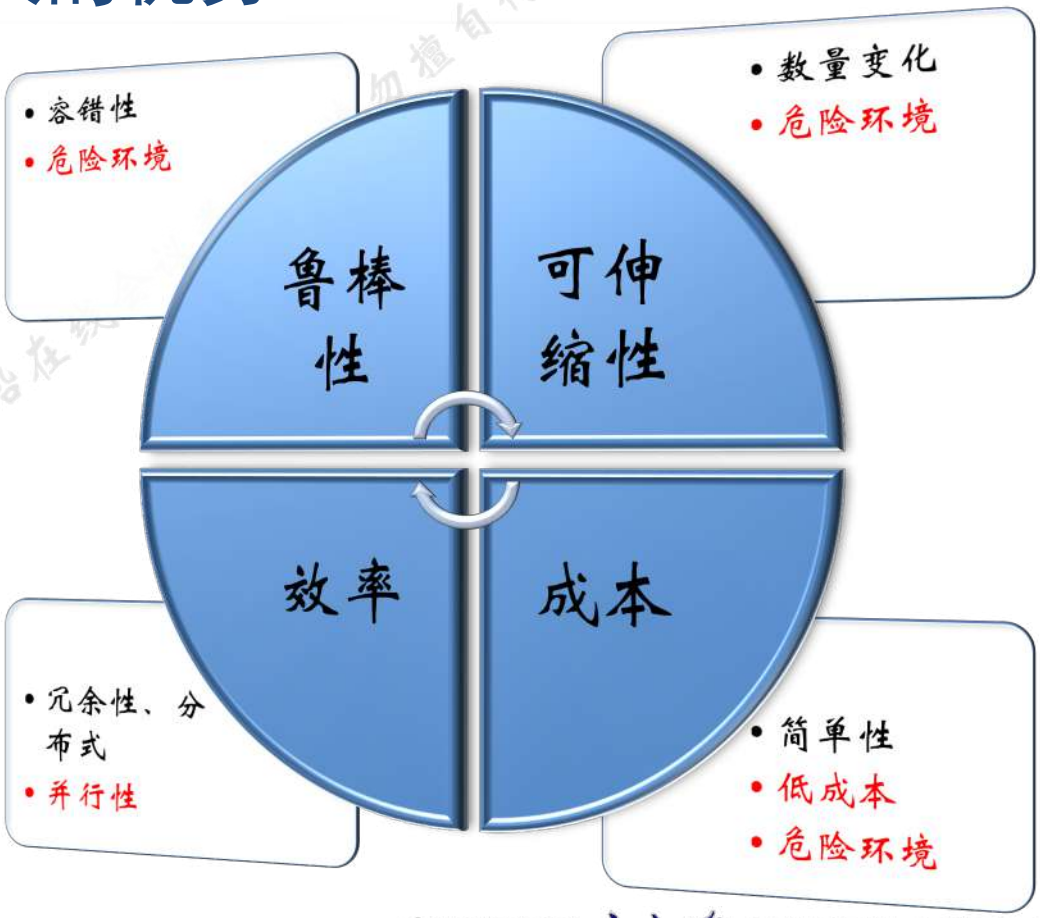




模拟仿真平台

多目标搜索问题与群体机器人的优势

- 现实任务
  - 灾后救援
  - 海难救援
  - 敌方潜艇搜寻
  - 战场目标摧毁
- 任务特征
  - 多个目标
  - 大空间
  - 效率需求
  - 可伴有危险



多目标搜索问题的假设

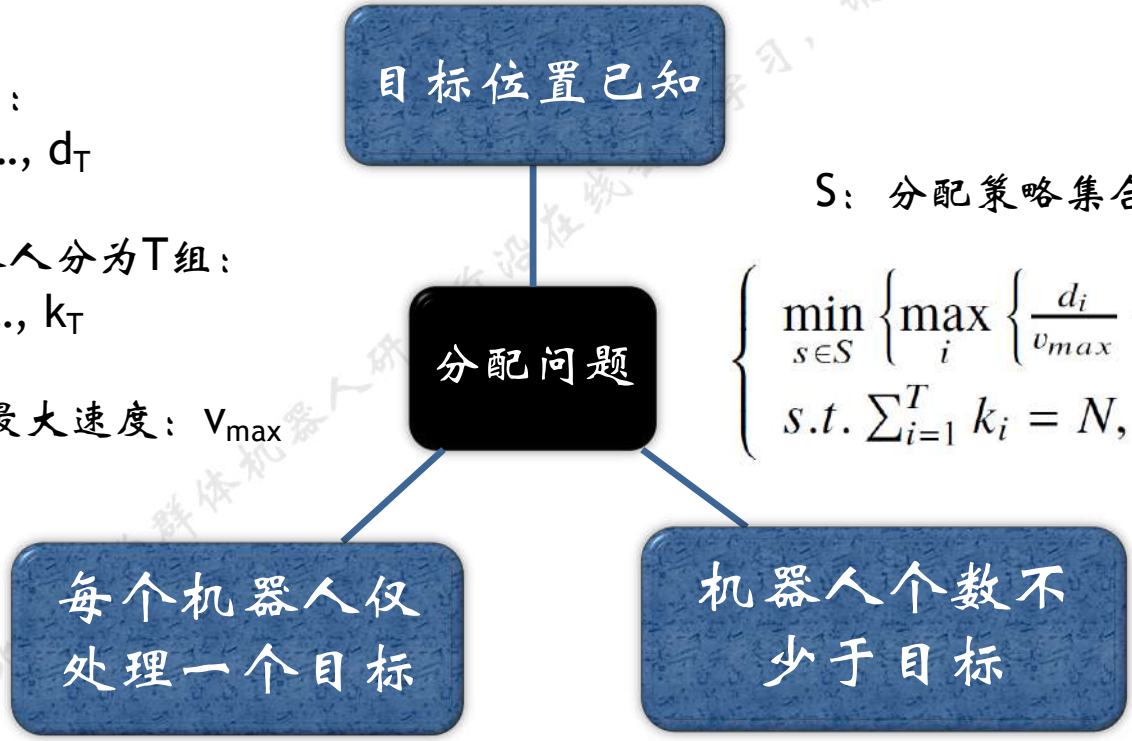
环境	目标	机器人
• 空间巨大	• 静止 • 小尺寸 • 大影响范围 • 适应度随距离递减 • 随机分布	• 无先验 • 局部交互 • 速度、存储有限 • 无集中控制 • 同一区域

多目标搜索问题的近似数学模型

T个目标：
$d_1, d_2, \ldots, d_T$

N个机器人分为T组：
$k_1, k_2, \ldots, k_T$

机器人最大速度：v_{max}



S：分配策略集合

$$\begin{cases} \min\limits_{s\in S}\left\{\max\limits_{i}\left\{\frac{d_i}{v_{max}}+\frac{10}{k_i}\right\}\right\} \\ s.t.\ \sum_{i=1}^{T} k_i = N, k_i \in N^+ \end{cases}$$

搜索效率的理论上界

近似计算

- 等面积圆形区域
- [108,117]

蒙特卡罗模拟

- 考虑阵列间隔（如200个机器人的阵列）
- [109,118]

$$\min_{s\in S}\left\{\max_{i}\left\{\frac{d_i}{v_{max}}+\frac{10}{k_i}\right\}\right\}$$

$$\approx E\left[\frac{d_T}{v_{max}}\right]+5.5\pm 4.5$$

默认参数配置：

- 地图1000*1000
- 10个目标
- 50个机器人

搜索策略性能的衡量指标

- 效率
  - 迭代次数的均值
- 稳定性
  - 迭代次数的标准差
- 规模敏感性
  - 增加一个机器人少需的迭代次数
- 并行处理能力
  - 增加一个目标多需的迭代次数
- 协作处理能力
  - 增加一次目标收集次数多需的迭代

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- **基于分组爆炸的多目标搜索策略**
- 基于三角编队搜索的多目标搜索策略
- 基于概率有限状态机的多目标搜索策略
- 基于深度学习和进化计算的多目标搜索策略
- 群体协作在游戏AI中的应用
- 总结与展望

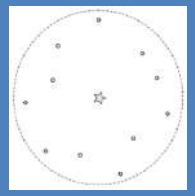
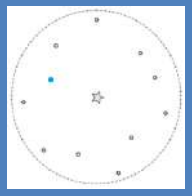
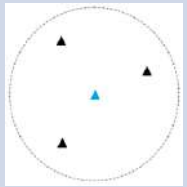
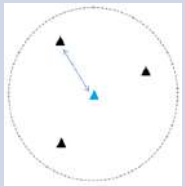
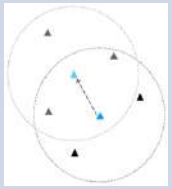
分组爆炸算法（Group Explosion Strategy, GES）

- 适应值分布密集
  - 只考虑细化搜索阶段和目标处理阶段

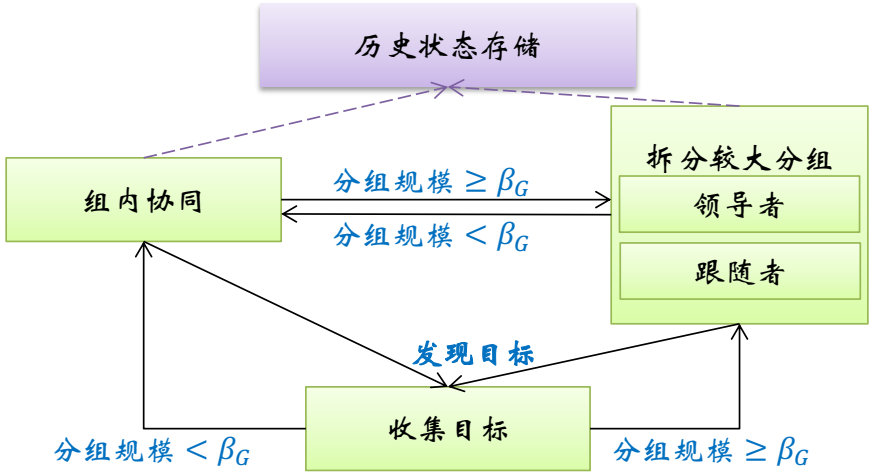
- 核心思想
  - 将群体分成多个小组
  - 组内协同加快搜索
  - 组间并行搜索多个目标
  - 分组大小阈值β_G

分组爆炸算法（Group Explosion Strategy, GES）

- 引入烟花爆炸机制

爆炸机制	 爆炸碎片	 爆炸半径	 新的爆炸中心
GES	 个体及其邻域	 邻域半径	 子群移动

- 算法流程



分组爆炸算法（Group Explosion Strategy, GES）

- 组内协同
  - 分组中心=>组内最优位置
    - $C_i(t) = \frac{\sum_{j \in N_i(t)} P_j(t)}{\|N_i(t)\|}$
  - 调整爆炸幅度
    - $G_i(t) = \left(P_b(t) - C_i(t)\right) * R_s$
    - R_s从$1-\beta_S$、 1和$1+\beta_S$中随机选择

- 拆分较大分组
  - 适应值最好的两个个体=>领导者
    - 相互排斥
    - $V_R(L_i) = \left(P_{L_i}(t) - P_{L_{1-i}}(t)\right) * \beta_R, i = 0,1$
  - 跟随者
    - 根据权值随机选择跟随一个领导者
      - $w(L_i) = F(L_i) + 1$
      - $P(L_i) = \frac{w(L_i)}{w(L_0)+w(L_1)}$
    - $G_i(t) = V_R(l_i) + \left(P_{l_i}(t) - P_i(t)\right) * R_s$

分组爆炸算法（Group Explosion Strategy, GES）

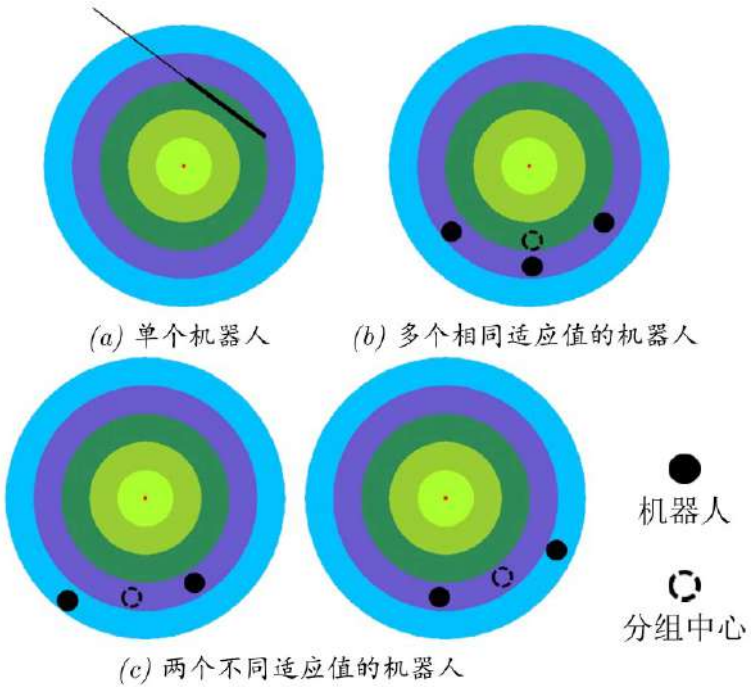
- 历史信息
  - $H_i(t) = (P_i(t) - h_i) * r$
  - 随机系数r服从$U[0.4,0.8]$

- 速度更新

$G_i(t)$	$H_i(t)$	$V_i(t)$
$\|G_i(t)\| > 0$	-	$G_i(t) + H_i(t)$
$\|G_i(t)\| = 0$	$\|H_i(t)\| > 0$	$H_i(t) + R_p$
	$\|H_i(t)\| = 0$	$V_i(t-1)$

改进的分组爆炸算法（IGES）

- 分组爆炸算法的不足之处



- 简化策略
  - 多个个体的分组
    - 只考虑临近个体信息
    - 简化拆分机制
  - 单个个体的分组
    - 只考虑历史记录信息

- 提高组内协同性能
  - 细化处理情况
    - 提高策略针对性
    - 简化单一策略
  - 分组所需个体减少
    - 减小分组阈值
    - 提高并行性

改进的分组爆炸算法（IGES）

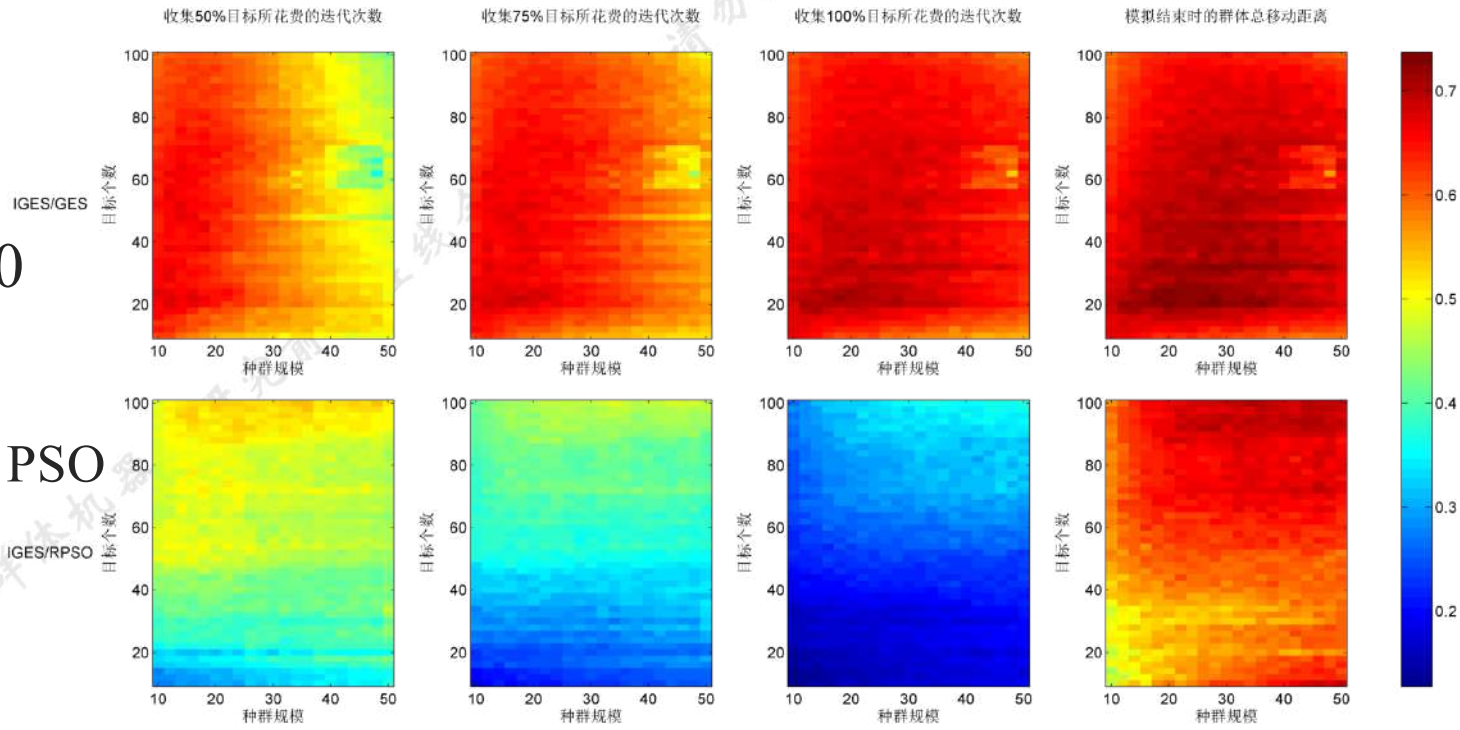
- 改进后的策略

群组大小	多个个体		单个个体		
适应值	适应值不同	适应值相同	历史最优	比上一代差	其他
策略$S_i(t)$	群体中心=>最优个体中心	远离群体中心	方向不变	向历史最优中心移动	
随机量r	$1/10$		0	1	$1/10$
简化	取消爆炸幅度调整		取消历史信息随机系数		

$$V_i(t) = S_i(t) + r \cdot R_p$$

实验结果和讨论

- 目标数量：10-100
- 个体数量：10-50
- 地图大小：1000*1000
- 对比算法RPSO
  - 基于典型群体智能算法PSO
  - 使用局部交互



实验结果和讨论

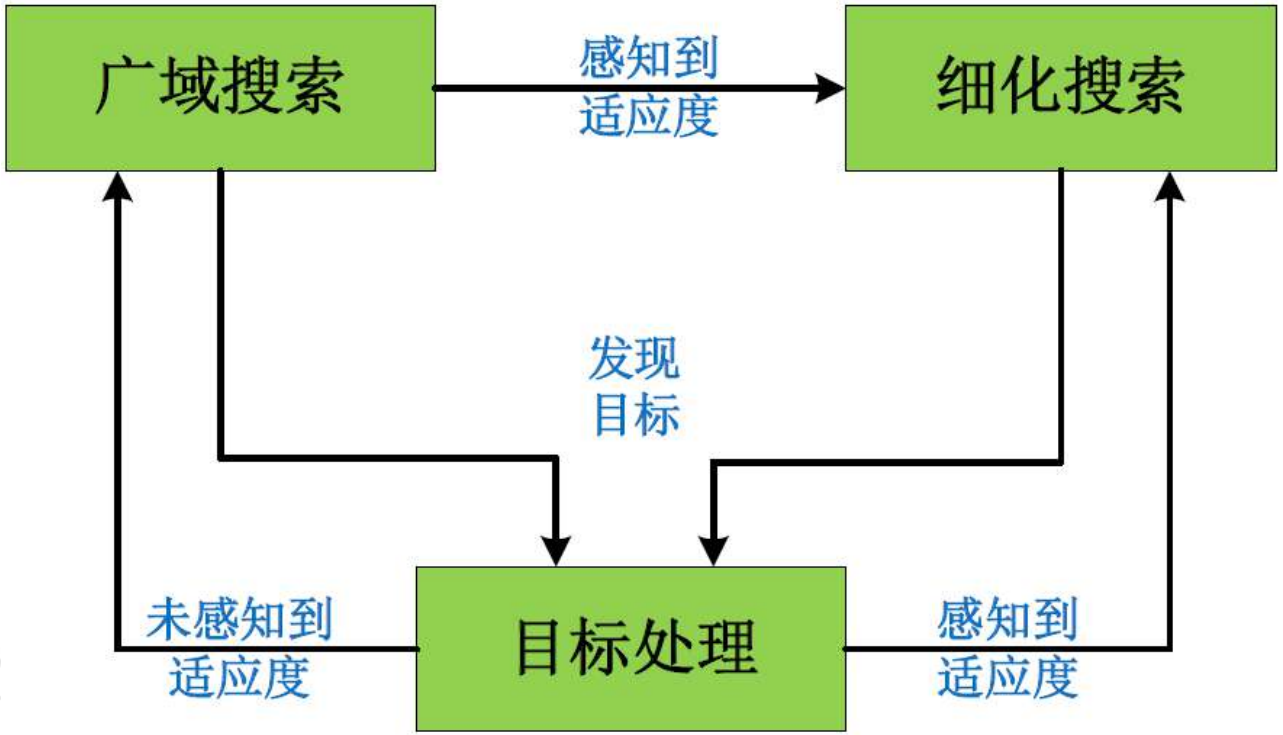
- GES算法
  - 大部分情况下优于RPSO
  - 个体比较密集时性能有所下降
- IGES算法
  - 策略简单
  - 计算复杂度低
  - 适应能力能力强
  - 相较于现有算法具有30%以上的提升

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- **基于三角编队搜索的多目标搜索策略**
- 基于概率有限状态机的多目标搜索策略
- 基于深度学习和进化计算的多目标搜索策略
- 群体协作在游戏AI中的应用
- 总结与展望

三角编队搜索（Triangle Formation Search, TFS）

- 三阶段搜索框架



三角编队搜索（Triangle Formation Search, TFS）

- 研究思路
  - 增强开采
    - 划分为固定小队
    - 队内整合信息
    - 小队间无协作
  - 增强探索
    - 借鉴随机搜索
    - 个体独立搜索
    - 个体间无协作

- 三角编队搜索策略的五个阶段
  - 初始分组：三机编队（领队和队员）
  - 初始扩散：邻居较少的方向
  - 无适应度区域搜索：随机搜索
  - 有适应度区域搜索：局部梯度估计
  - 目标收集：广播目标位置

三角编队搜索（Triangle Formation Search, TFS）

- 关键技术

三角编队搜索（Triangle Formation Search, TFS）

- 三角梯度估计

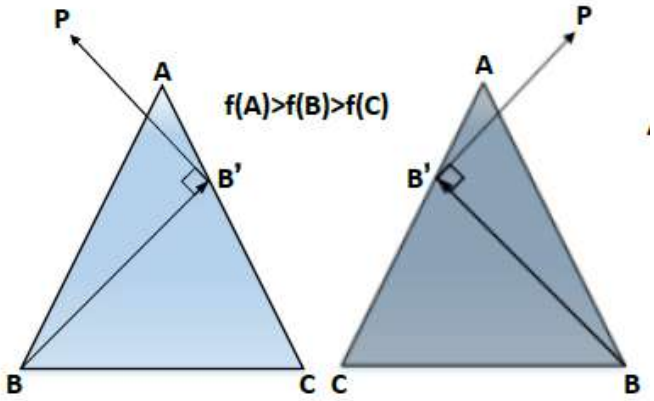
  —局部线性变化，等值线的法向量

Algorithm 1 三角梯度估计

Require: $P_i(x_i, y_i), f_i(i = 1, 2, 3)$: 三个点的位置和适应度值。
$\vec{v}_{last}$: 机器人r的上一速度

Ensure: $\vec{v}$: 机器人r的新速度

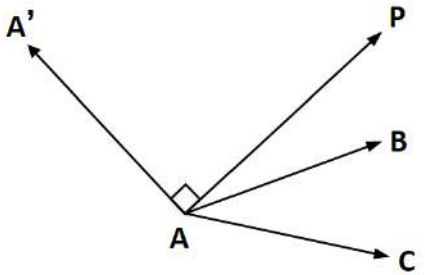
1: $(P_A, P_B, P_C) \leftarrow sort(P_1, P_2, P_3)$ by $f_i(i = 1, 2, 3)$，使得$f_A \geq f_B \geq f_C$
2: **if** $f_A = f_C$ **then**
3: $\vec{v} \leftarrow \vec{v}_{last}$ {情况 I}
4: **else if** $f_A = f_B$ **then**
5: $\vec{v} \leftarrow (P_A + P_B)/2 - P_C$ {情况 II}
6: **else if** $f_B = f_C$ **then**
7: $\vec{v} \leftarrow P_A - (P_B + P_C)/2$ {情况 III}
8: **else**
9: $P_{B'} \leftarrow P_A + (P_C - P_A) \times (f_A - f_B)/(f_A - f_C)$ {情况 IV}
10: $BB' \leftarrow P_{B'} - P_B$
11: 由公式(3.2)计算$B'P$
12: $\vec{v} \leftarrow B'P$
13: **end if**

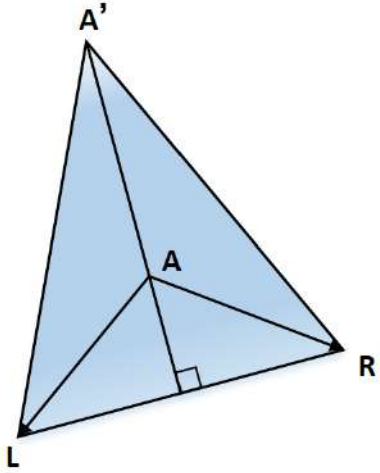


$$\begin{cases} \vec{B'P} \cdot \vec{BB'} = 0 \\ \vec{B'P} \cdot \vec{B'A} > 0 \end{cases}$$

三角编队搜索（Triangle Formation Search, TFS）

- 角色切换与队形控制
  - 角色切换
    - 减少剧烈转向
    - 最优个体为领队
    - 领队广播位置
    - 队员确定角色
  - 队形控制
    - 队员确定位置
    - 领队监测距离





三角编队搜索（Triangle Formation Search, TFS）

- 独立搜索策略

  —随机搜索
  - 列维飞行
  - 弹道移动
  - 间歇式搜索

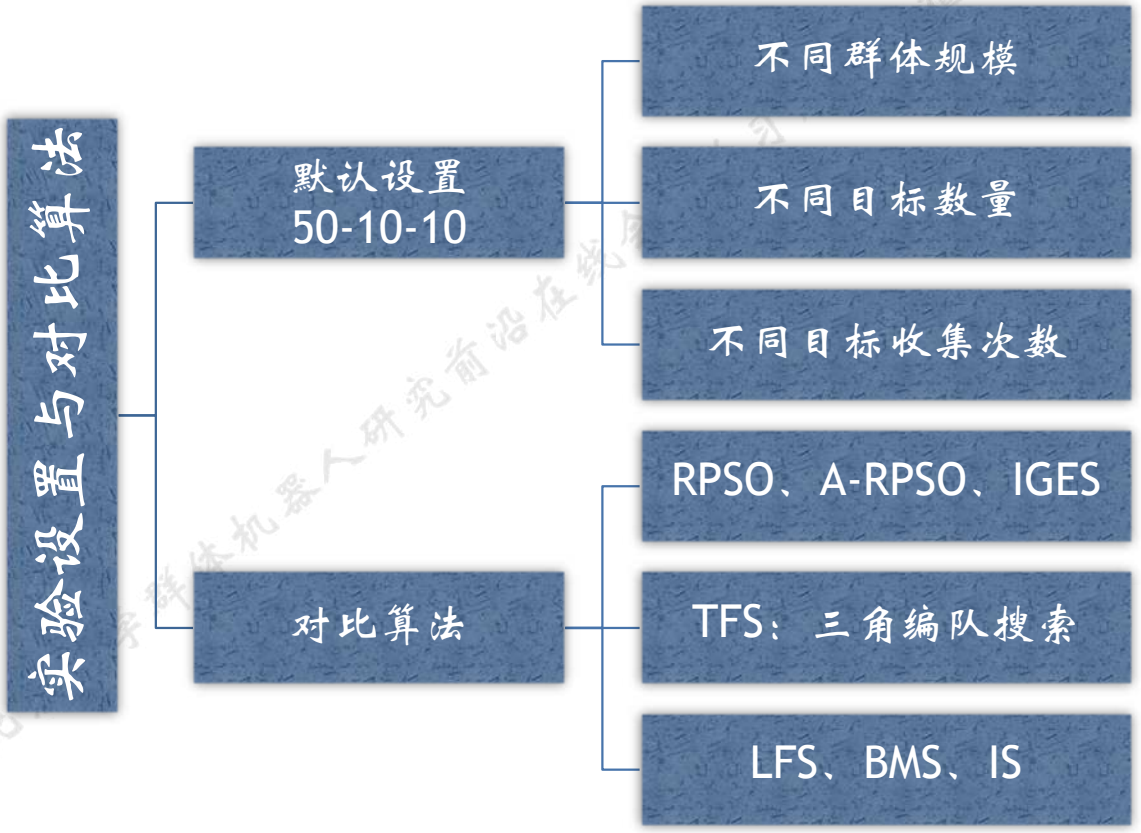
  —三角梯度估计
  - 当前位置
  - 历史最优
  - 历史最差

  —惯性机制
  - 稳定方向
  - 跳出极值
  - $v_{t+1} = wv_t + (1-w)v_{s,t}$

- 三种随机搜索

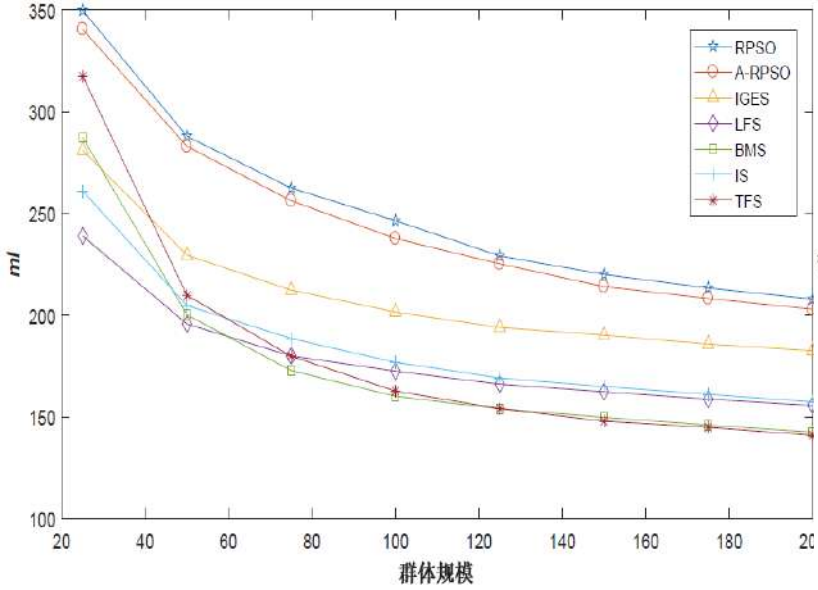
列维飞行	弹道移动	间歇式搜索
• 幂律分布 • u=1.001 • LFS	• 直线运动 • BMS	• 双状态 • 指数分布 • IS

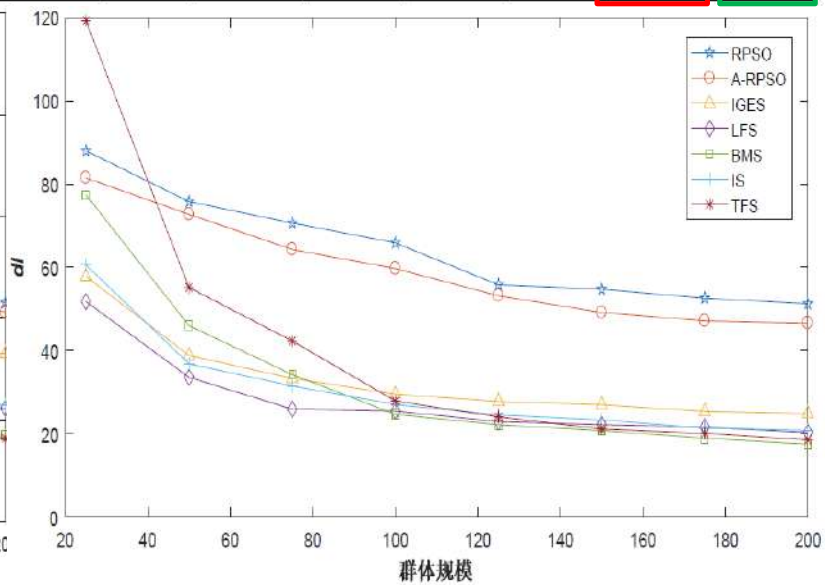
实验设置与对比算法



不同群体规模

群体规模	*RPSO*		*A – RPSO*		*IGES*		*LFS*		*BMS*		*IS*		*TFS*	
	mI	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*
25	349.79	87.96	340.54	81.45	281.07	57.80	**238.98**	**51.76**	287.30	77.37	261.02	60.55	317.45	119.2
50	287.98	75.76	283.05	72.70	229.53	38.83	**195.71**	**33.58**	200.26	45.97	204.78	36.82	209.79	55.14
75	262.56	70.64	256.27	64.28	212.38	33.35	180.04	**25.96**	**172.85**	34.26	188.59	31.53	179.90	42.30
100	246.49	65.87	237.73	59.70	201.60	29.53	172.51	25.48	**160.21**	**24.70**	176.87	27.06	162.75	27.92
125	229.18	55.78	225.28	53.17	194.04	27.79	166.19	22.96	**153.96**	**22.21**	169.20	24.54	154.09	24.09
150	220.07	54.74	214.11	49.15	190.21	27.02	162.40	22.24	149.84	**20.83**	164.99	23.37	**147.99**	21.22
175	213.43	52.58	208.26	47.14	185.87	25.41	158.77	21.64	146.09	**18.99**	161.17	21.43	**144.92**	20.15
200	207.94	51.24	202.96	46.58	182.63	24.89	155.68	20.38	142.55	**17.58**	157.40	20.87	**141.01**	18.57





小结

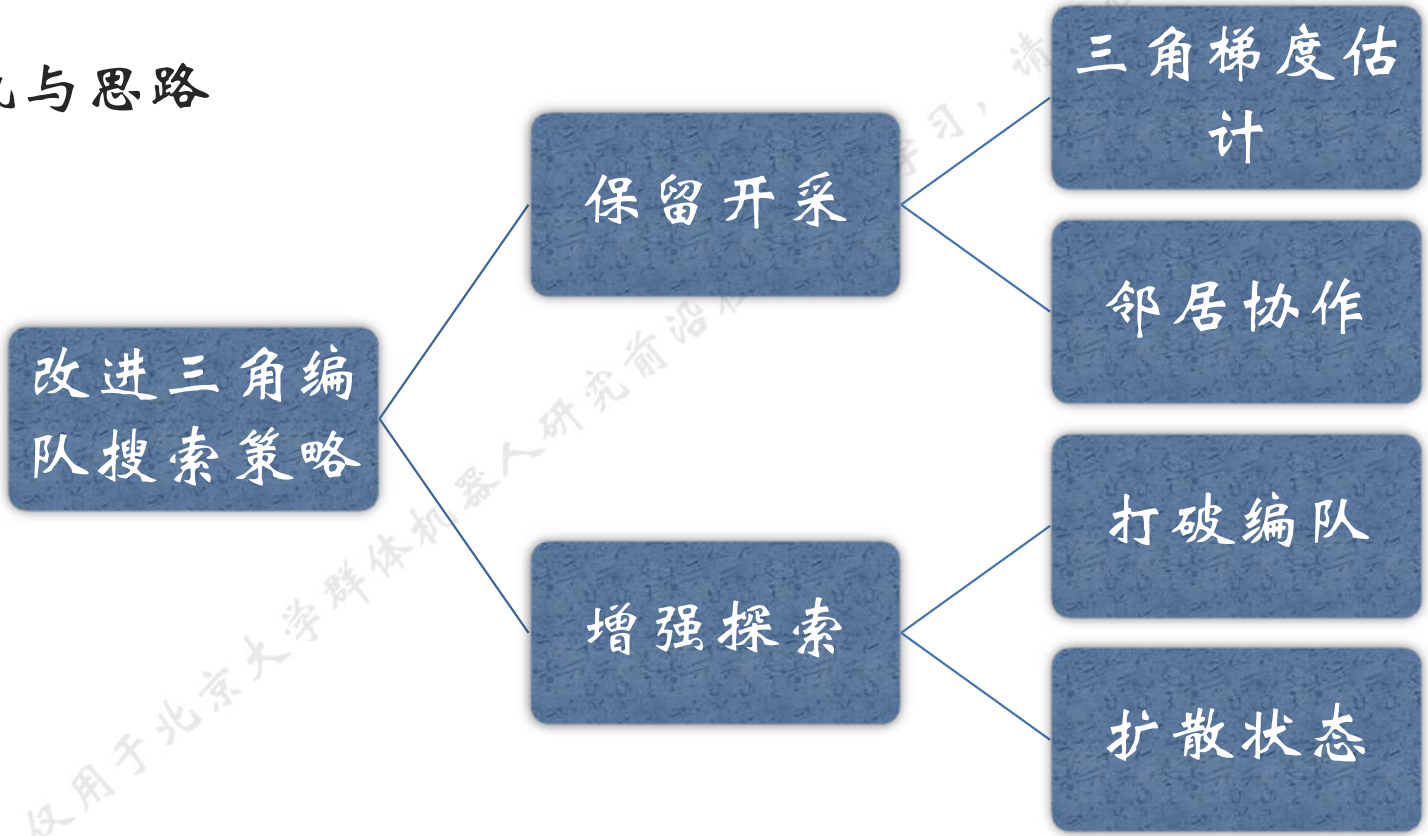
衡量指标	三角编队搜索策略TFS	独立搜索策略
效率和稳定性	大规模下优 小规模下差	基本较优 尤其小规模
规模敏感性	高	低（BMS较高）
并行处理能力	较强	强
协作处理能力	强	弱
启发	开采能力很强 探索能力较强，但资源仍相对过于集中 打破固定编队	探索能力很强 需加强协作以增强开采

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- 基于三角编队搜索的多目标搜索策略
- **基于概率有限状态机的多目标搜索策略**
- 基于深度学习和进化计算的多目标搜索策略
- 群体协作在游戏AI中的应用
- 总结与展望

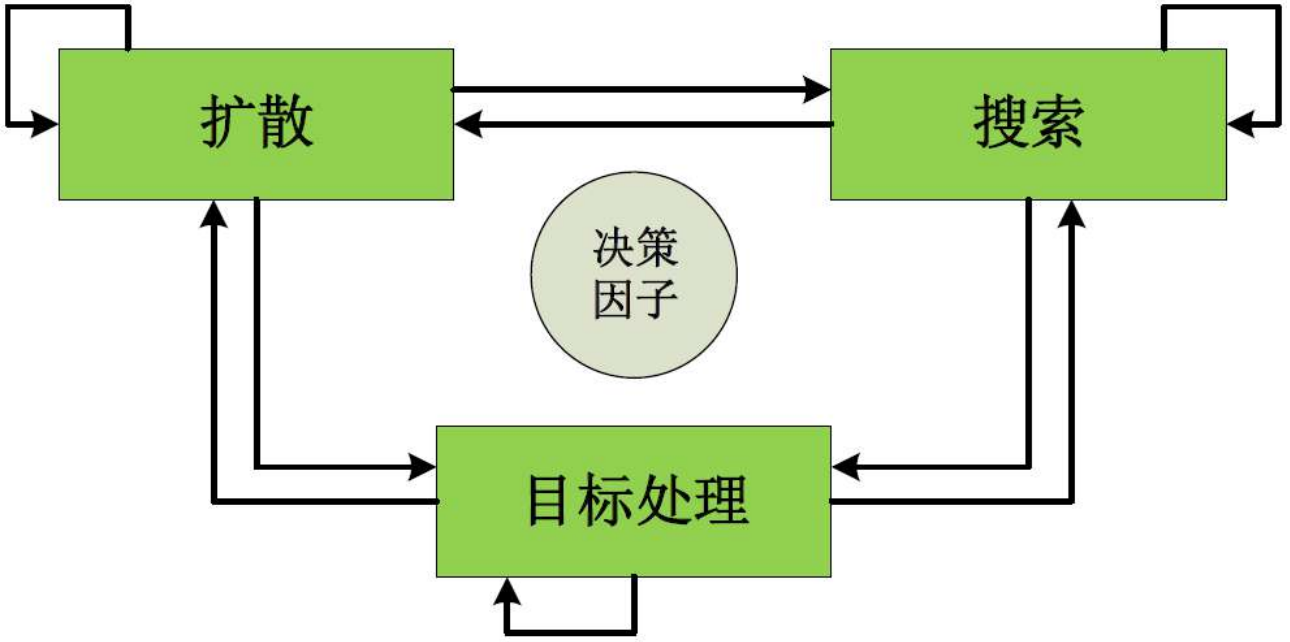
概率有限状态机(Probabilistic Finite State Machine, PFSM)

- 研究动机与思路



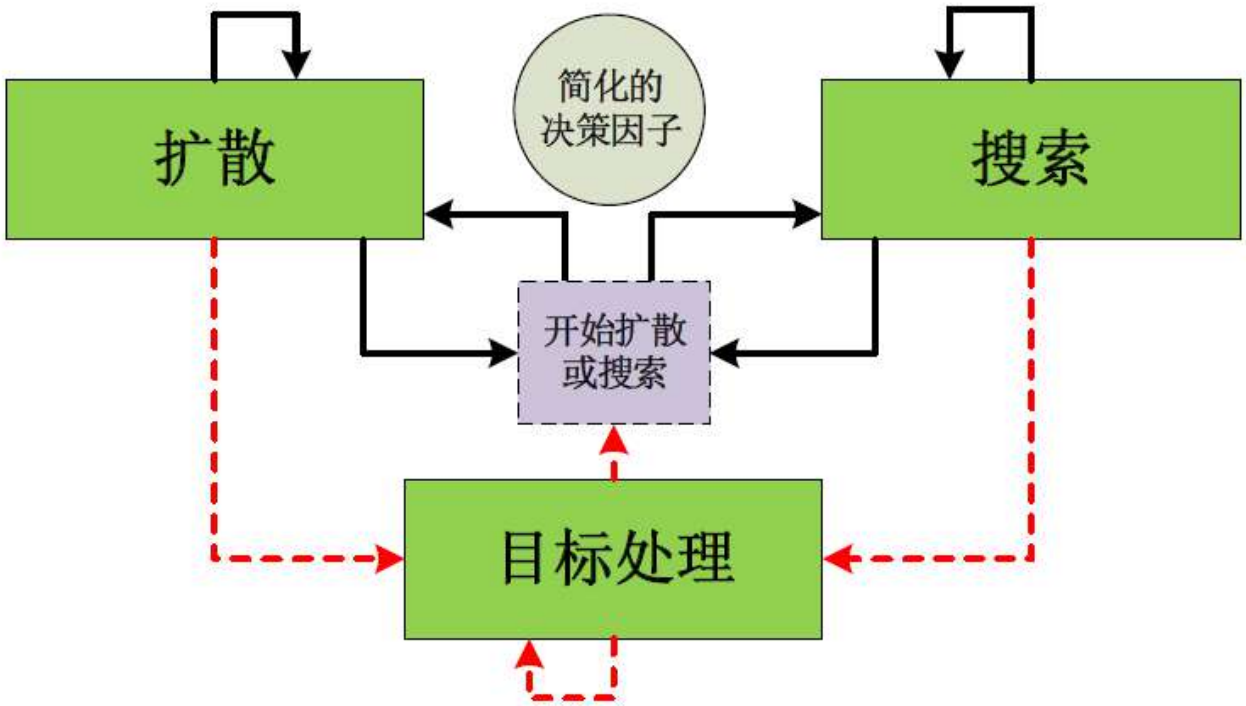
概率有限状态机(Probabilistic Finite State Machine, PFSM)

- 三状态概率有限状态机



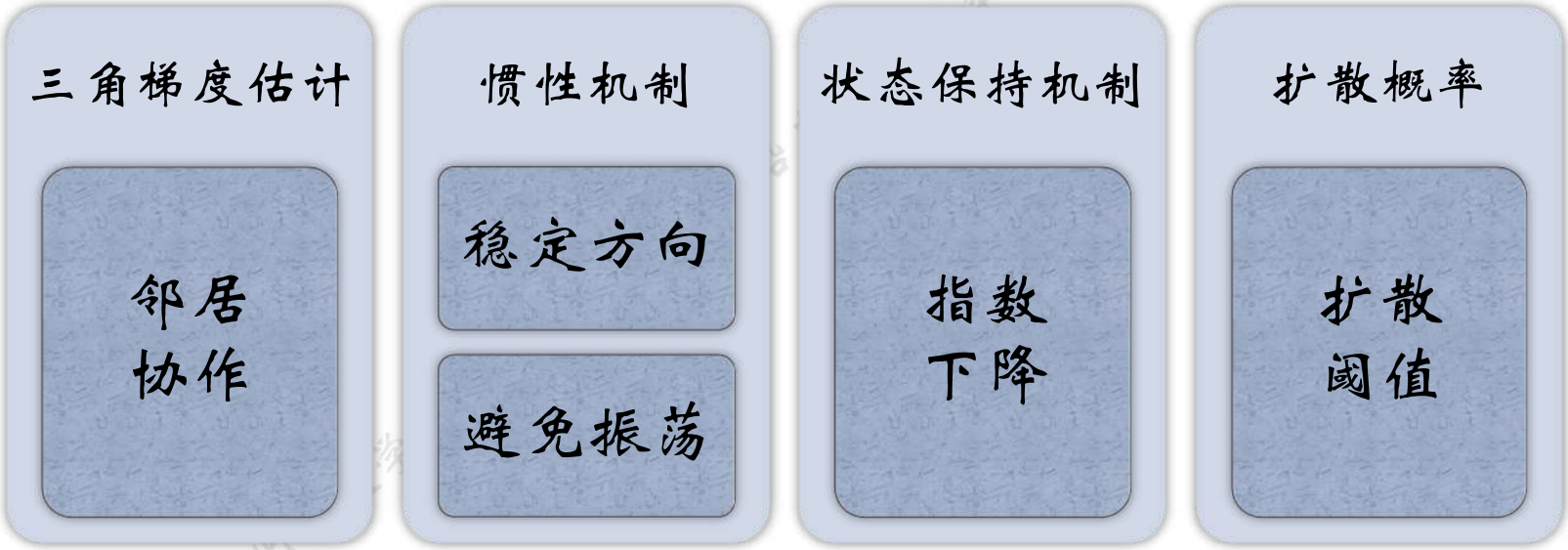
概率有限状态机(Probabilistic Finite State Machine, PFSM)

- 简化的概率有限状态机



概率有限状态机(Probabilistic Finite State Machine, PFSM)

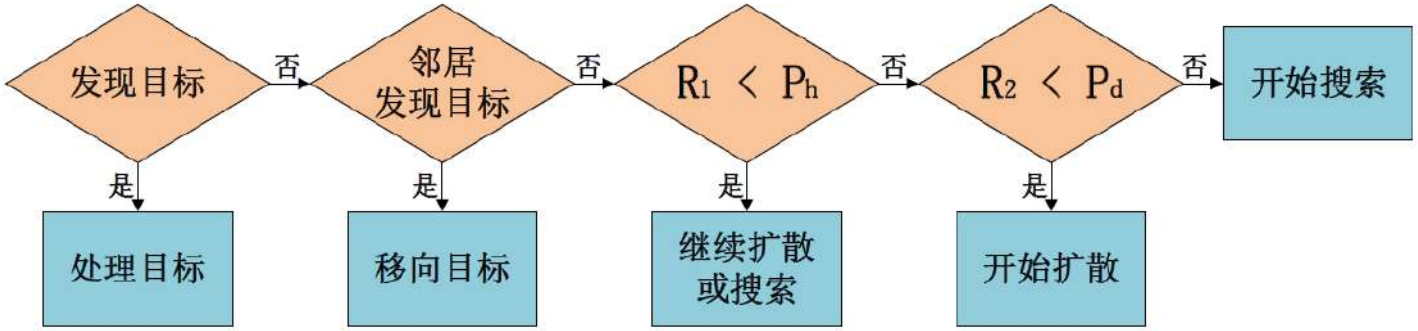
- 关键技术



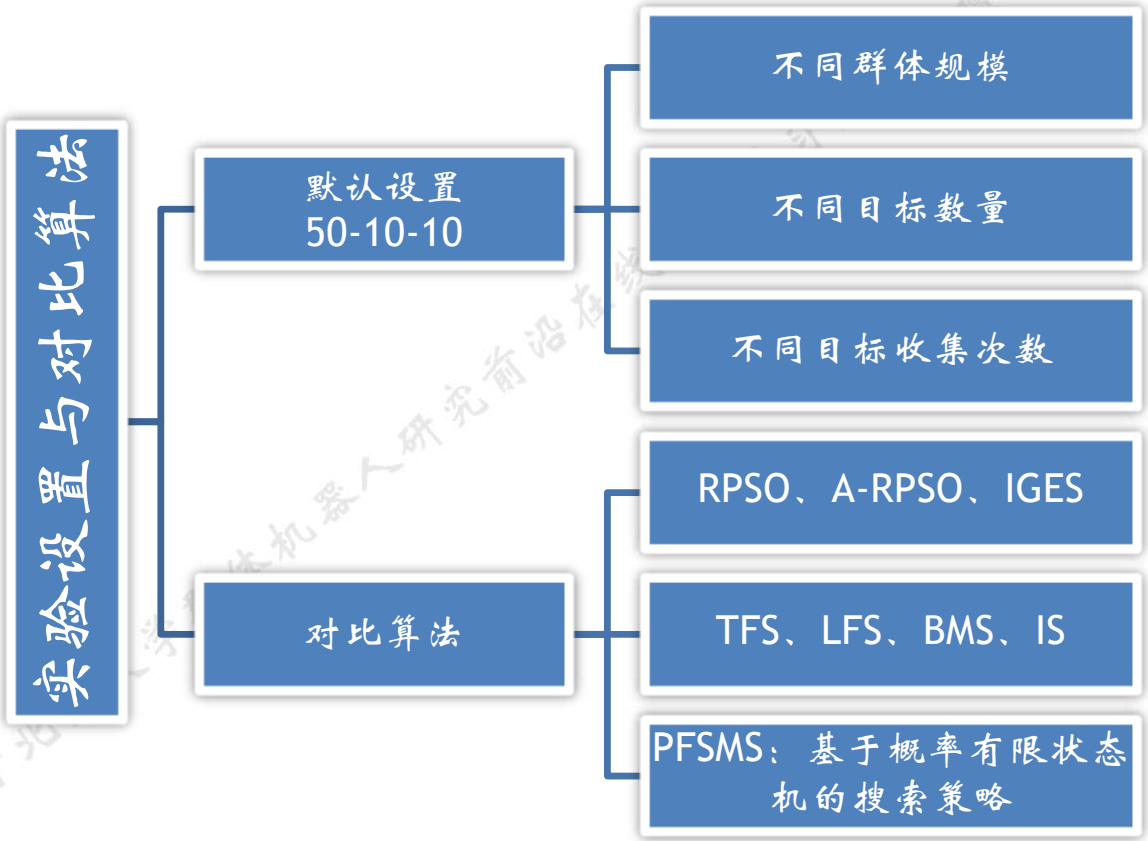
$$P_h = P_{ini}^{N_h} \qquad P_d = \begin{cases} 1 - \frac{T_b}{N_b} & , N_b > T_b \\ 0 & , N_b \le T_b \end{cases}$$

概率有限状态机(Probabilistic Finite State Machine, PFSM)

- 个体决策流



实验设置与对比算法



不同问题配置下各对比策略的效率排名

	RPSO	*A – RPSO*	*IGES*	*LFS*	*BMS*	*IS*	*TFS*	*PFSMS*
群体规模	8.0	7.0	5.8	3.5	2.9	4.6	3.3	1.0
目标数量	7.6	7.1	6.3	2.0	3.0	4.0	5.0	1.0
目标收集次数	8.0	7.0	6.0	2.0	3.4	4.4	4.3	1.0
总体	7.9	7.0	6.0	2.5	3.1	4.3	4.2	1.0

- 群体规模50
- 目标数量10
- 目标收集次数10

小结

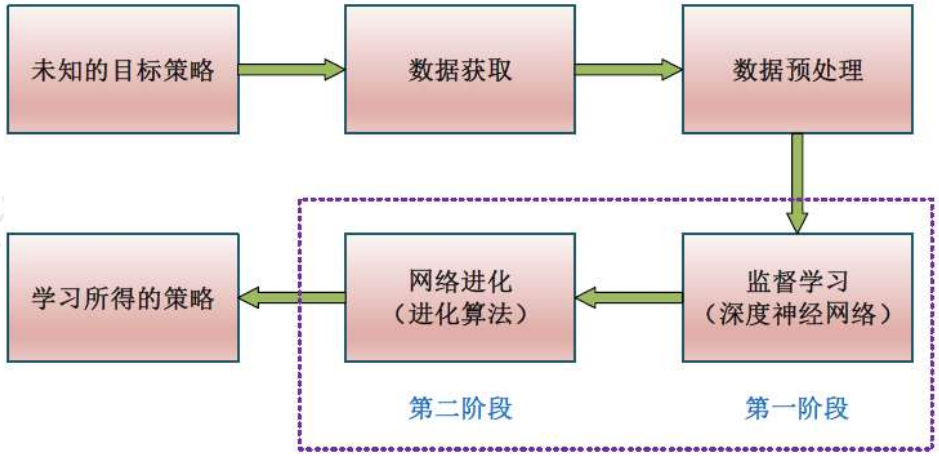
- 几乎在各指标上都显示压倒性优势
- 大群体规模下效率接近理论最优值
- 有效地平衡了群体的探索与开采

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- 基于三角编队搜索的多目标搜索策略
- 基于概率有限状态机的多目标搜索策略
- **基于深度学习和进化计算的多目标搜索策略**
- 群体协作在游戏AI中的应用
- 总结与展望

基于深度学习和进化计算的多目标搜索策略

- 研究动机
  - 一生物群体的自组织行为
    - 蚂蚁觅食
    - 鸟类迁徙
    - 具体策略未知（目标策略）
  - 一学习生物群体的行为
    - 采集数据
    - 策略设计
    - 机理揭示
  - 一优化模型
    - 进化算法
    - 强化学习
- 基于深度学习和进化计算的两阶段学习框架



基于深度学习和进化计算的多目标搜索策略

- 目标策略——PFSMS
- 30维输入，2维输出

索引	描述
0	当前状态，三种取值，-1表示搜索，1表示扩散，0表示目标处理
1	当前状态已持续的迭代次数
2-3	当前的速度
4-6	邻居是否发现目标以及目标的位置
7-11	邻居总个数，上、下、左、右四个方向的邻居个数
12	历史记录个数
13-15	历史上的最优位置及其适应度值
16-18	历史上的最差位置及其适应度值
19	当前适应度值
20-22	最优邻居的位置及其适应度值
23-25	最差邻居的位置及其适应度值
26-29	四维随机数信息，用于概率性决策，可在行为观测时同步产生

索引	描述
0	新的移动方向（角度值），取值范围为$(-\pi, \pi]$
1	新的状态

基于深度学习和进化计算的多目标搜索策略

- 数据采集和预处理

独立同分布

- 地图种子
- 迭代次数
- 机器人编号

数据均衡

- 搜索状态 1/5->1/2
- 150万条
- 3:1:1

归一化

- 训练集
- 均值
- 标准差

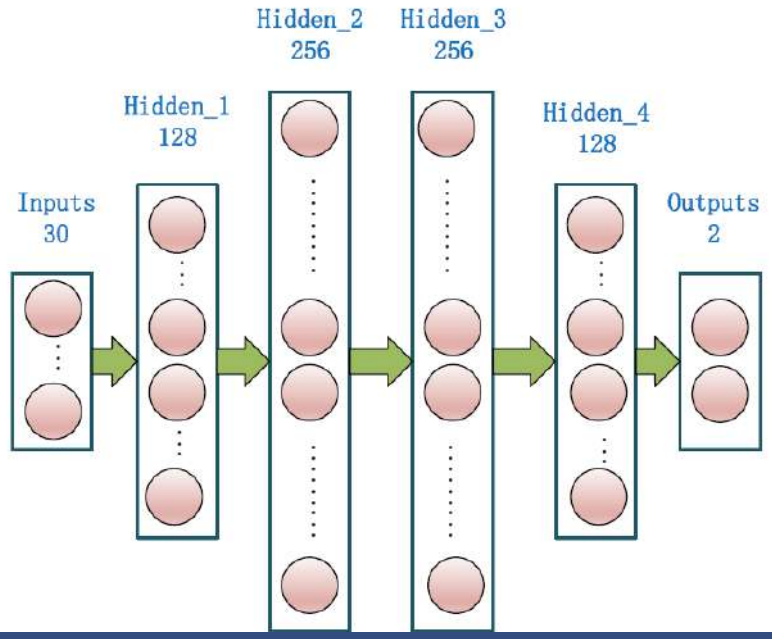
$$x^i = \frac{x^i - \mu^i}{s^i}$$

基于深度学习和进化计算的多目标搜索策略

- 修正线性单元 ReLU
- 比例指数线性单元 SeLU
- 栈式自编码器 SAE，dropout

$$\text{rectifier}(x) = \max(0, x)$$

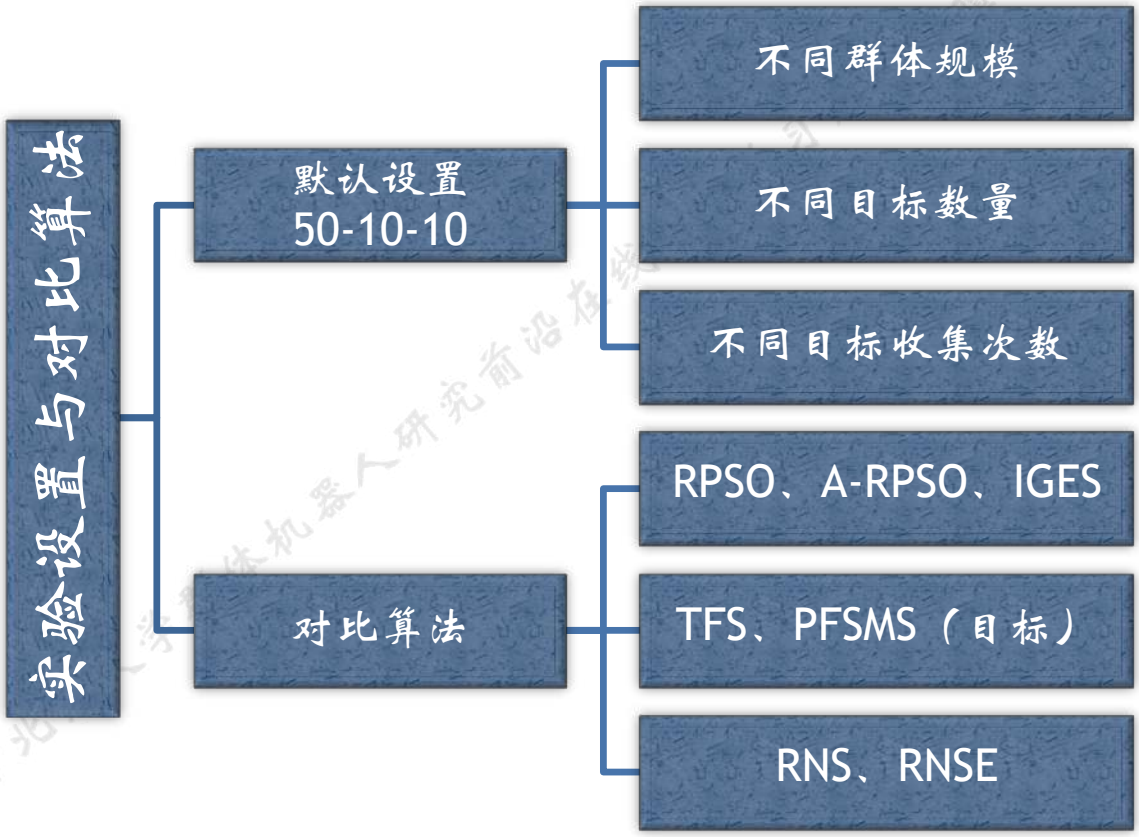
$$selu(x) = \lambda \begin{cases} x & , if \ x > 0 \\ \alpha e^x - \alpha, & if \ x \le 0 \end{cases}$$



基于进化计算的策略学习

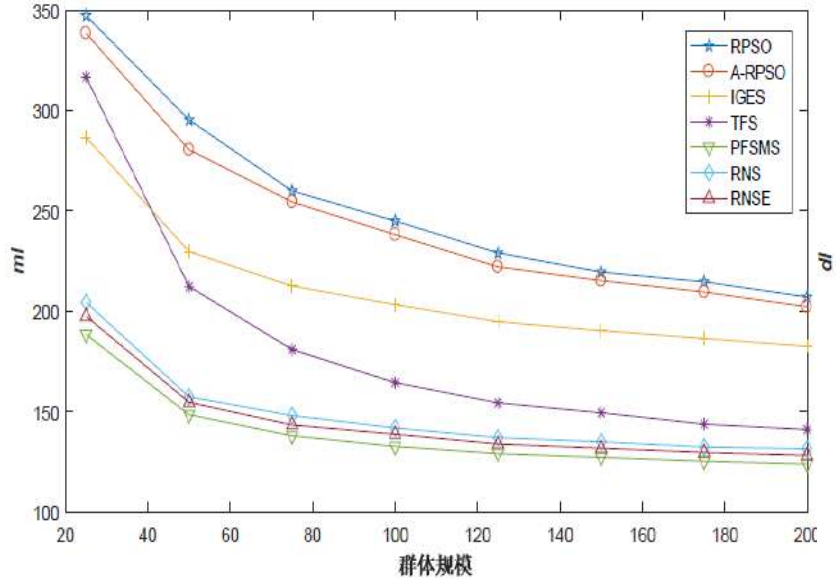
$$\sigma' = \frac{\sigma}{\sqrt{1000}} < 1$$

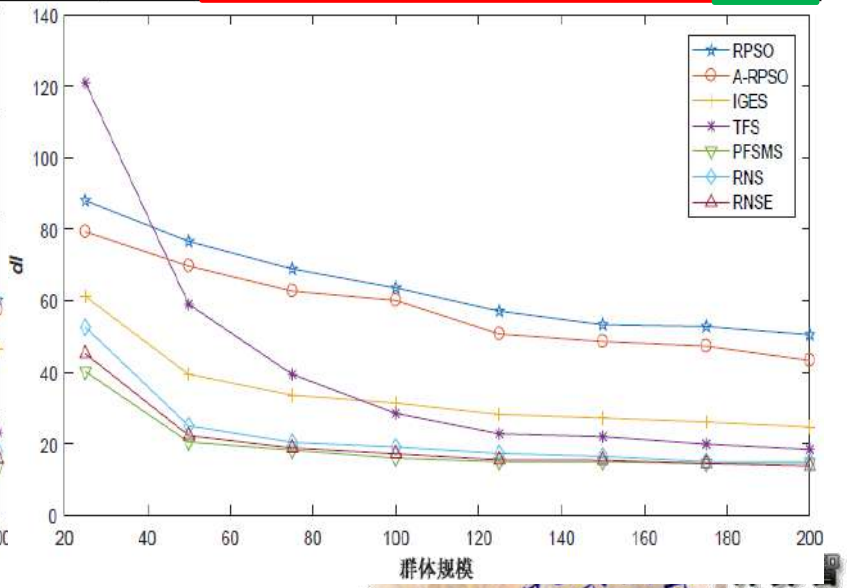
实验设置与对比算法



不同群体规模

群体规模	RPSO		A－RPSO		IGES		TFS		PFSMS		RNS		RNSE	
	mI	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*	*mI*	*dI*
25	347.4	88.0	338.6	79.3	286.7	61.2	316.7	121.0	188.4	40.2	204.4	52.6	197.5	45.2
50	295.4	76.6	280.5	69.7	229.7	39.4	212.3	59.0	148.4	20.5	157.3	25.0	154.6	22.3
75	260.0	68.9	254.5	62.7	212.6	33.6	180.9	39.4	137.9	18.2	148.0	20.4	143.3	18.8
100	245.0	63.6	238.2	60.1	203.3	31.4	164.4	28.5	132.6	16.0	141.8	19.1	138.7	17.2
125	229.1	57.1	222.1	50.7	194.8	28.2	154.3	22.8	129.0	15.0	137.0	17.3	133.8	15.5
150	219.5	53.3	215.3	48.6	190.3	27.2	149.4	22.0	127.1	15.0	134.9	16.5	131.7	15.4
175	214.7	52.8	209.6	47.3	186.4	26.1	143.7	19.9	125.2	14.5	132.3	15.0	129.6	14.5
200	207.1	50.5	202.2	43.3	182.5	24.7	141.0	18.4	123.8	14.7	131.4	15.0	128.1	13.8





小结

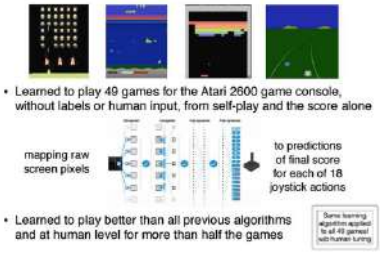
- 在各项指标上接近目标策略
- 经过进化学习的策略表现更优
- 验证了两阶段学习框架的有效性
- 网络结构和进化算法的设置有借鉴意义

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- 基于三角编队搜索的多目标搜索策略
- 基于概率有限状态机的多目标搜索策略
- 基于深度学习和进化计算的多目标搜索策略
- **群体协作在游戏AI中的应用**
- 总结与展望

群体协作在游戏AI中的应用

1997年，IBM的深蓝战胜了当时的国际象棋特级大师 Garry Kasparov



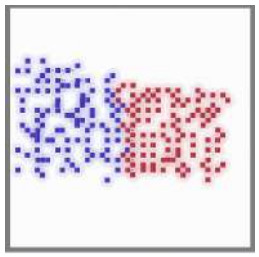
2015年，Google-DeepMind发表文章称深度强化学习在Atari游戏任务上超越人类表现

2017年，Google-DeepMind的Alphago围棋程序战胜柯洁

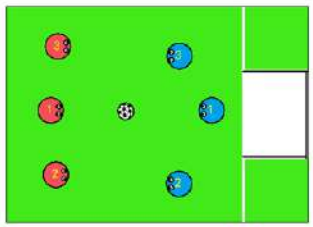
2017年，OpenAI的Five机器人，在标准比赛规则下，在刀塔2的 1v1比赛中与世界冠军邓迪（Dendi）进行对抗。

基于群体协作的游戏AI？

Pommerman
(炸弹人)



Battle Game



Soccer Game

StarCraft
(星际)

基于注意力机制的多智能体强化学习状态表示方法

- 研究动机

实现全局协同仍需信息集中式的处理模式

多智能体拓扑连接的动态性和规模变化给特征表示带来了极大的不确定性

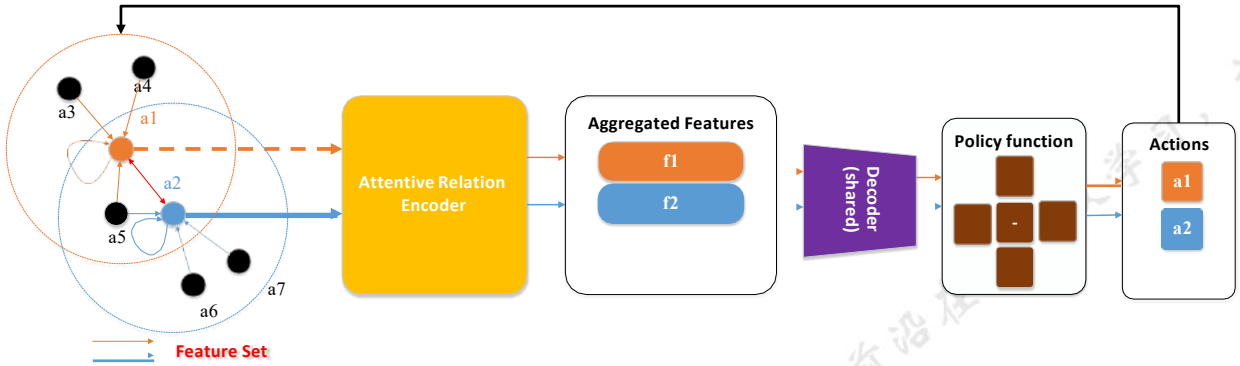
传统方法缺乏灵活性

特征并联→维度灾难；离散化/平均化→精度损失

不同邻居在不同时刻的效用也不同

问题：如何为智能体构建一个稳定且高效的特征表示？

图视角下的多智能体状态表示学习



多智能体拓扑关系存在潜在的图结构

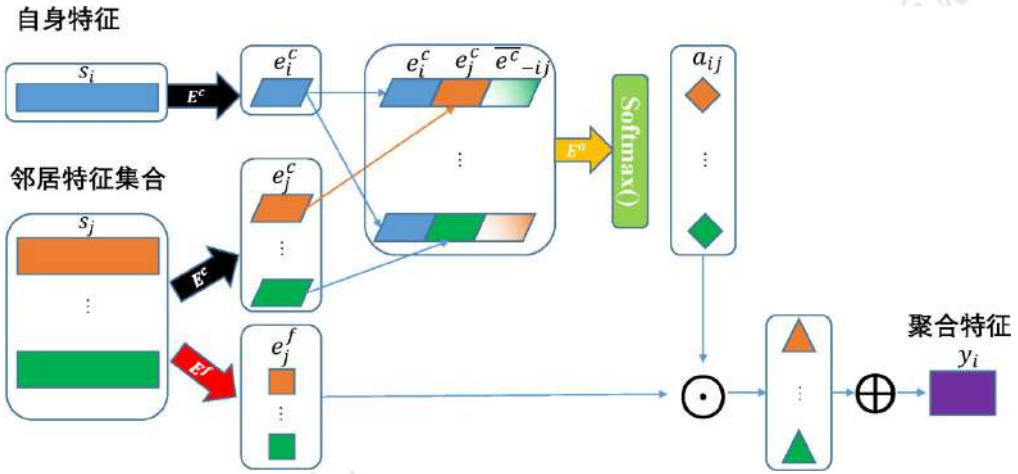
- 智能体为节点（node）
- 智能体状态为节点特征（feature of node）
- 邻居之间的连接为边（edge）
- 邻居之间的交互关系为（有向）边特征（feature of edge）
- 时变关系动态图$G_t = (A, E_t)$

图注意力机制 + 多智能体深度强化学习

北京大学 计算智能实验室 Computational Intelligence Laboratory, Peking University

基于注意力机制的多智能体强化学习状态表示方法

- 注意力关联编码器
  - 注意力关系编码模型（Attentive Relational Encoder, ARE）



$$e_i^f = E^f(s_i), e_i^c = E^c(s_i)$$

$$e_{ij}^a = E^a\left(e_i^c, e_j^c, \overline{e^c}_{-ij}\right)$$

$$a_{ij} = \frac{\exp(e_{ij}^a)}{\sum_{k \in G_i} \exp(e_{ik}^a)}$$

$$y_i = \sum_j a_{ij} e_j^f$$

$$\pi_i = decoder(y_i)$$

基于注意力机制的多智能体强化学习状态表示方法

排列不变性

- Softmax操作对排列置换函数不敏感

数量不变性

- 子图规模不固定，输出的状态表示固定（使用了Sum-pooling）

计算效率

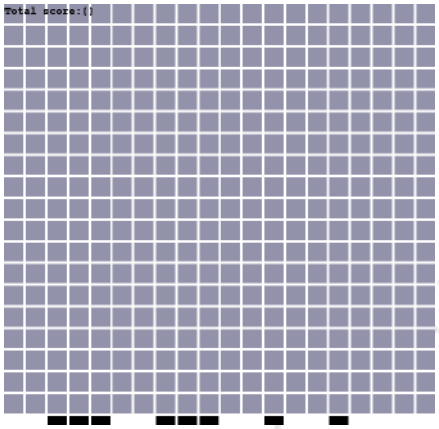
- 所有操作均可以并行化，且所有执行模块共享

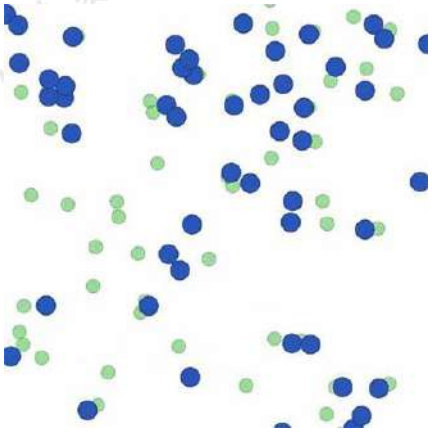
效用区分能力

- 注意力机制逐个计算邻居的重要性权重

实验结果和讨论——实验任务

- 任务





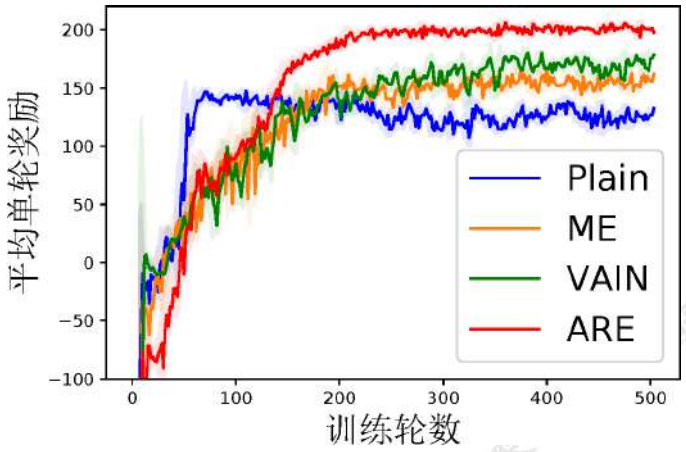
接球任务

- 根据视野中的目标分布调整位置
- 实现“互惠共利”

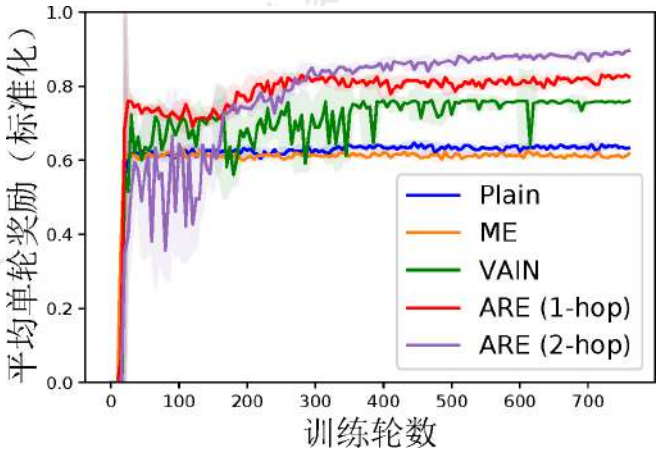
目标覆盖任务

- 就近覆盖地标
- 躲避其他智能体行进路线

实验结果和讨论——训练收敛性验证



8-智能体接球任务



50-智能体目标覆盖任务

对邻居个体的差异化建模能显著提高个体表示学习的能力

复杂交互场景建模多轮聚合能显著提高性能

实验结果和讨论——协同指标验证

对比方法	评价指标		
	目标收集率	p_1	p_2
独立编码（Plain）	76.4%	0.701	0.201
平均嵌入（ME）	85.1%	0.831	0.149
交互网络（VAIN）	87.0%	0.841	0.133
注意关联编码器（ARE）	**90.6%**	**0.989**	**0.128**

多智能体接球任务协同性指标

对比方法	奖励值	碰撞次数	地标覆盖比例
独立编码（Plain）	0.63	5.02	92.5%
单跳-平均嵌入（1-hop ME）	0.62	**0.06**	36.4%
双跳-平均嵌入（2-hop ME）	0.65	**0.05**	36.3%
单跳-交互网络（1-hop VAIN）	0.71	1.82	98.1%
双跳-交互网络（2-hop VAIN）	0.76	1.80	97.2%
单跳-注意关联编码器（1-hop ARE）	**0.83**	1.86	**98.5%**
双跳-注意关联编码器（2-hop ARE）	**0.90**	**1.20**	**98.3%**

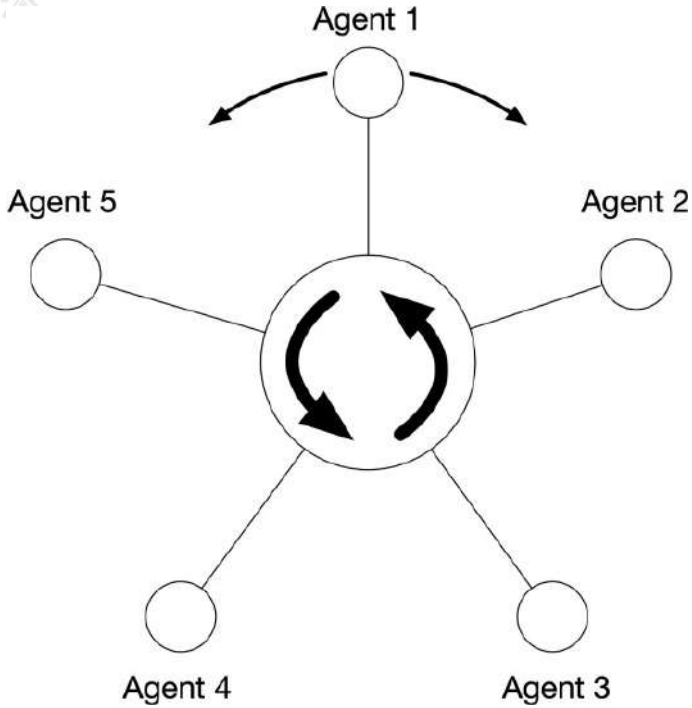
多智能体目标覆盖任务协同性指标

小结

- 提出一种基于注意力机制的关系编码器用于多智能体表示学习
- 满足动态交互环境中特征表示的排列不变性/规模不变性/效用区分能力
- 性能优于特征并联/平均池化等特征表示方法

基于协同隐空间的多智能体强化学习探索方法

- 强化学习方法依赖探索机制（试错）
- 多智能体学习探索难
  - 随着智能体数量增多，高效的协同模式更难挖掘（组合动作空间$O(|A|^N)$）
  - 协同模式可能存在多个局部最优
  - 某一个有效的动作需要其他智能体的配合才能体现出来

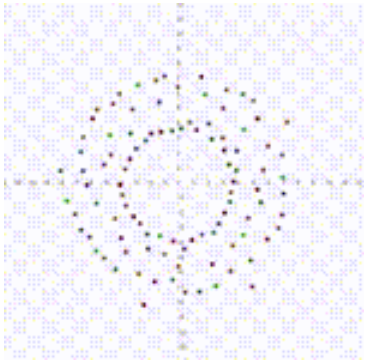


已有的探索方法

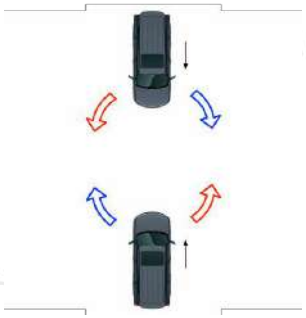
单智能体探索方法

扩展到多智能体系统时存在的问题

协同模式的观察



多机器人冲突消解



协同防撞策略

低维结构

- 协同模式存在低维流型
- “协同隐空间”→动作空间

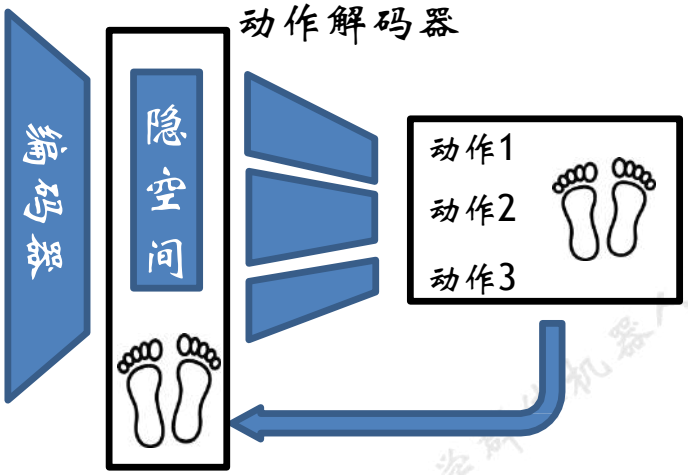
共享结构

- 共同知识
- 协同决策的一致性

多态结构

- 协同模式不唯一
- 协同隐空间存在多样性

隐空间探索方法

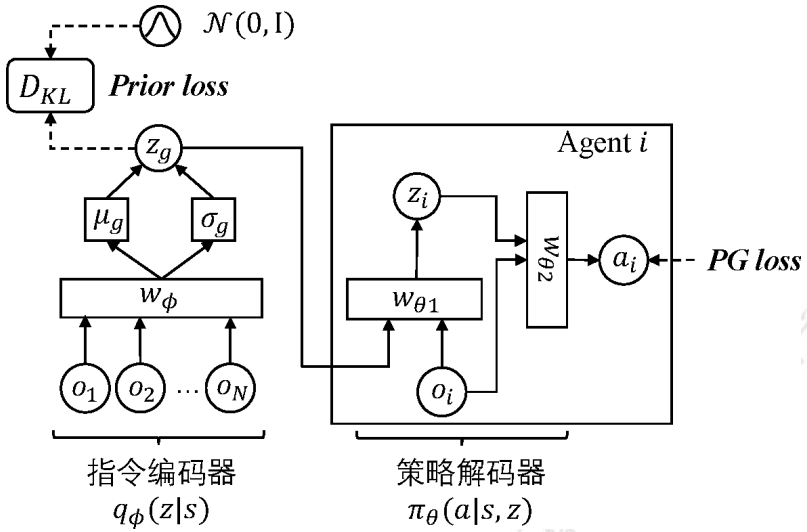


探索空间的转移：
原始联合动作空间→隐空间

隐空间保持多样性：
随机采样的形式

隐空间结构共享

基于低维协同模式的多智能体探索方法



隐空间探索多智能体强化学习的计算图
（Feudal Latent-space Exploration, FLE）

指令编码器$q_\phi(z|s)$

$$\mu_g(s) = W_\mu \phi(s) + b_\mu$$
$$log\sigma(s)^2 = W_\sigma \phi(s) + b_\sigma$$
$$z_g = \mu_g(s) + \sigma_g(s) \odot \epsilon$$

策略解码器$\pi_\theta(a|s,z)$

$$\pi_\theta^i(a_i|o_i, z_g) = \pi_{\theta_2}(a_i|o_i, \pi_{\theta_1}(o_i, z_g))$$

训练方法

在MADDPG的基础上加入隐空间编码器模型的训练

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}_{d\sim\mathcal{D}}\left[\begin{array}{c}\nabla_{\theta_i}\pi_{\theta_i}(o_i, z_g)\nabla_{a_i} Q_i^{\pi}(s, a_1, ..., a_n)|\\ a_i = \pi_{\theta_i}(o_i)\\ z_g = q_\phi(s)\end{array}\right]$$

$$\nabla_\phi J(\phi) =$$

$$\frac{1}{N}\sum_i \mathbb{E}_{d\sim\mathcal{D}}\left[\begin{array}{c}\nabla_\phi q_\phi(z_g|s)\nabla_{z_g}\pi_{\theta_i}(o_i, z_g)\nabla_{a_i} Q_i^{\pi}(s, a_1, ..., a_n)|\\ a_i = \pi_{\theta_i}(o_i)\\ z_g = q_\phi(s)\end{array}\right]$$

$$-\beta\nabla_\phi D_{KL}(q_\phi(z_g|s)||p(z)),$$

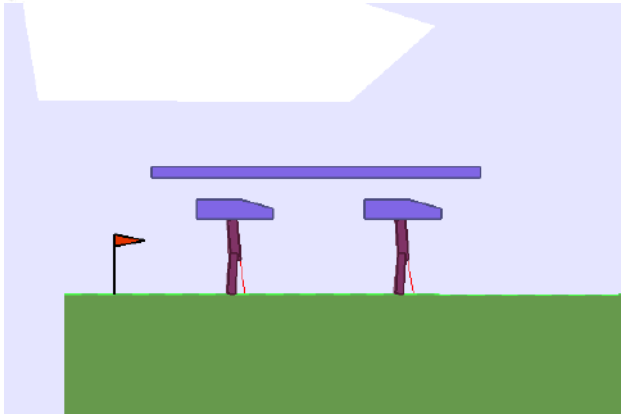
正则项

单位高斯先验

实验结果和讨论——实验任务

目标拾取任务（WaterWorld）

- 共同抓取目标（绿色）
- 躲避“毒药”（红色）



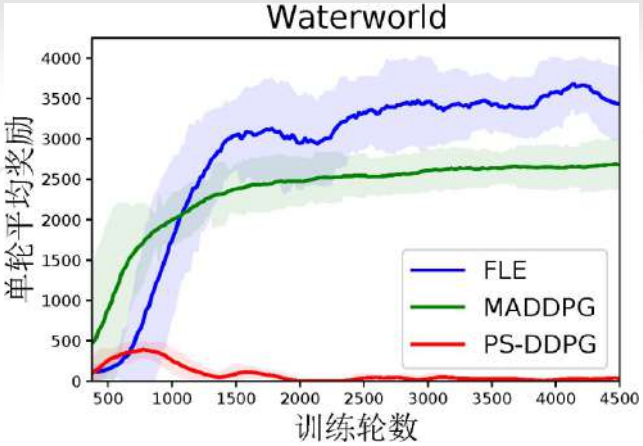
协同搬运任务（Multi-Walker）

- 学习行走策略的同时保持长棍的平衡
- 需要一致的行走节奏

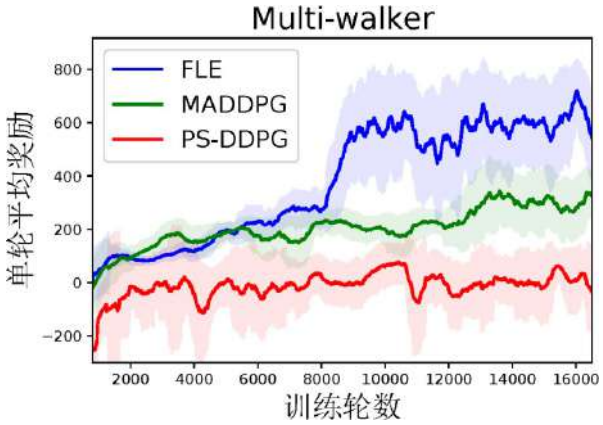
实验结果和讨论

FLE能收敛到更高的平均奖励

协同策略的多样性使得FLE的方差较大

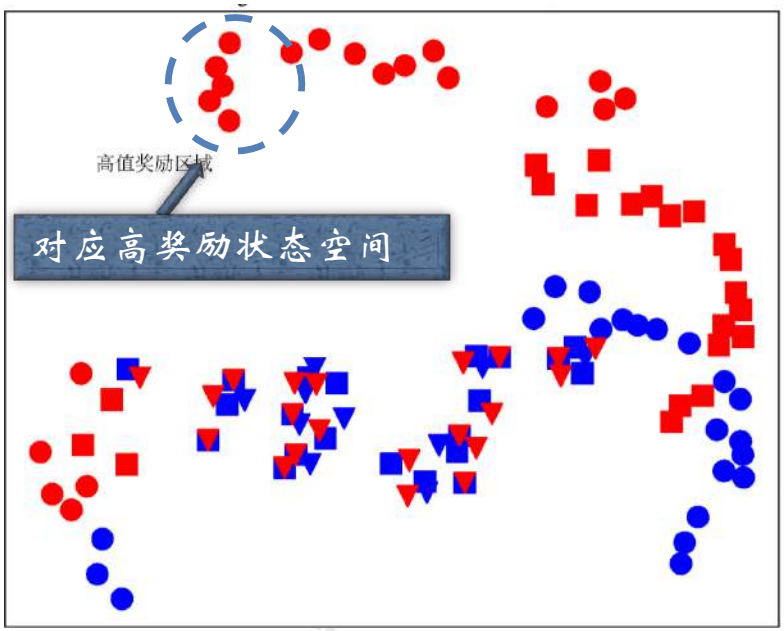


5-智能体目标拾取任务

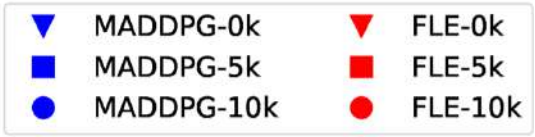


2-智能体协同搬运任务

实验结果和讨论——探索效率的可视化分析



轨迹数据降维可视化（t-sne）



FLE能快速摆脱初级阶段的探索，快速找到高奖励对应的状态空间

小结

- 提出一种隐空间探索用于学习低维协同模式
- 复杂协同任务中的探索效率优于朴素的随机噪声探索方法
- 可视化分析证明隐空间的学习可以掌握协同模式的多样性

应用研究：群体协作在游戏AI中的应用

- 星际争霸II

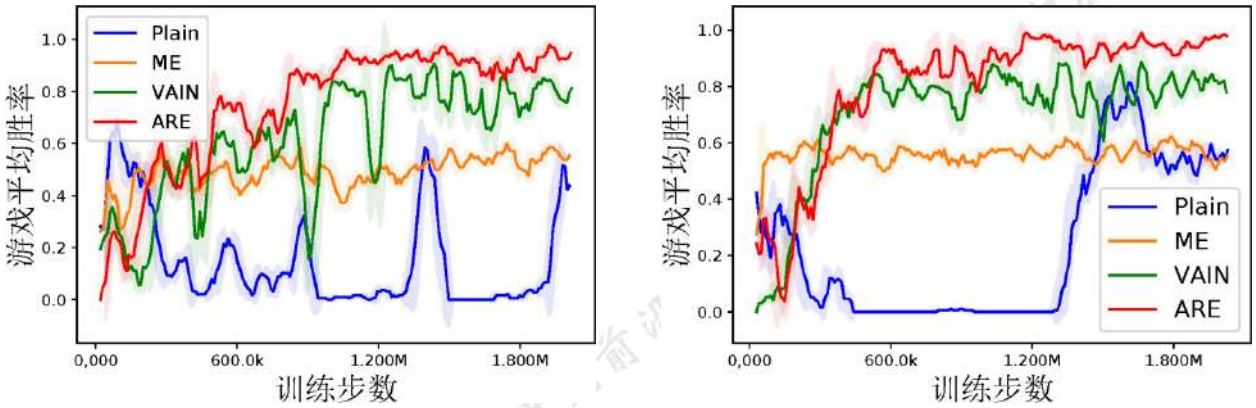
- 人类游戏史上最困难、也是最成功的游戏
- 多层次游戏机制
  - 建筑、科技、兵种、攻击策略、……
- 人工智能研究的“练兵场”
  - Google-DeepMind、Oxford等机构研究

应用研究：群体协作在游戏AI中的应用

- 实验场景：星际101对战场景（训练微操作）
- 智能体建模：
  - 局部特征：[move_features, enemy_feature, ally_feature, own_feature, obs_agent_id, obs_last_action]
  - 动作：[no_op, stop, north, south, west, east, attack/heal]
  - 奖励：(delta_enemy + delta_deaths - delta_ally + result) / max_reward * reward_scale_rate
  - 目标：消灭对方全部兵种单位，获得游戏胜利

9 Marines vs. 4 Roaches

实验结果与分析



Methods	9M vs. 4R 胜率	9M vs. 4R 奖励	10M vs. 10M 胜率	10M vs. 10M 奖励
独立编码（Plain）	46.4%	14.70	58.9%	13.91
平均嵌入（ME）	55.1%	15.31	56.8%	13.09
交互网络（VAIN）	83.0%	18.41	80.3%	15.22
注意关联编码器（ARE）	**96.6%**	**19.02**	**98.4%**	**19.13**

小结

- 将星际争霸2中的微操建模成多智能体协同任务
- 多智能体协同策略能以超过96%的胜率战胜非作弊难度的最强内置AI
- 展现出一些高级协同攻击技巧

大纲

- 研究背景与意义
- 群体机器人多目标搜索问题
- 基于分组爆炸的多目标搜索策略
- 基于三角编队搜索的多目标搜索策略
- 基于概率有限状态机的多目标搜索策略
- 基于深度学习和进化计算的多目标搜索策略
- 群体协作在游戏AI中的应用
- **总结与展望**

总结与展望

- 目前，群体智能研究正处于迅速发展阶段
- 新的自然/生物的机理和现象是发展新型群体智能算法和模型的源动力
- 群体智能应用正在许多领域逐步展开，尤其是群体机器人的研究和应用正在获得大家的广泛关注
- 广泛深入的实际应用，仍然面临下列挑战:
  - 高维大数据
  - 多目标
  - 不确定变化环境
  - 动态的协同协作
- 群体智能是通往强AI的方向，具有远大前景！

北京大学
PEKING UNIVERSITY

谢谢大家!

火龙
Beijing Haidian

Scan the QR Code to add me on WeChat

北京大学 计算智能实验室
Computational Intelligence Laboratory, Peking University