MAS-Bench: Parameter Optimization Benchmark for Multi-agent Crowd Simulation

Shusuke SHIGENAKA[1] Shunki TAKAMI[1] Shuhei WATANABE[2]
Yuki TANIGAKI[4] Yoshihiko OZAKI[1,3] Masaki ONISHI[4]
[1] University of Tsukuba [2] University of Freiburg [3] GREE, Inc.
[4] National Institute of Advanced Industrial Science and Technology (AIST)

https://github.com/MAS-Bench/MAS-Bench

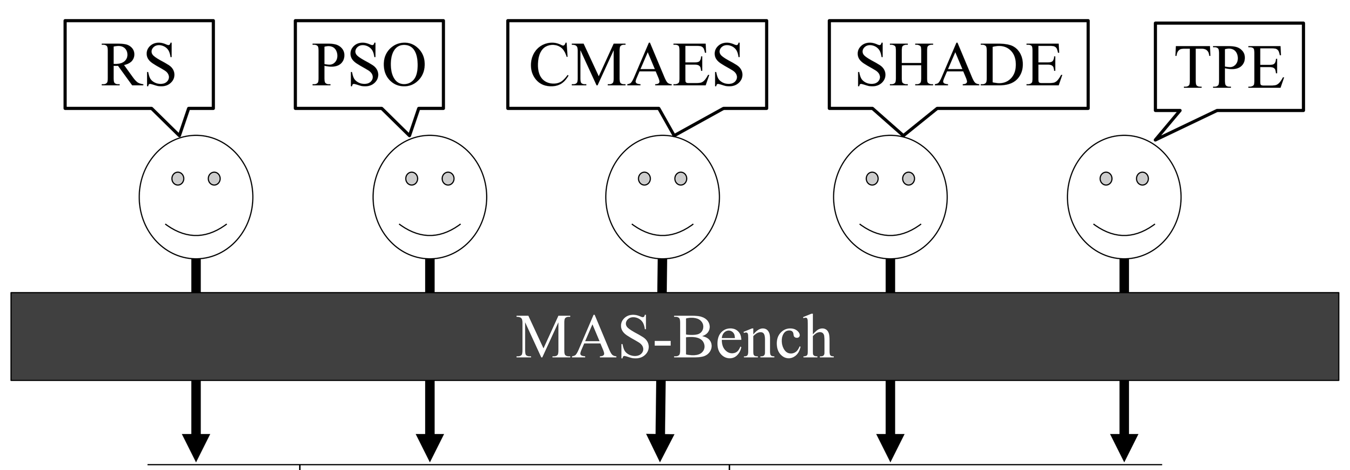
Introduction

Why don't you try your method on our MAS-Bench?
MAS-Bench is a benchmark to evaluate parameter optimization methods used in Multi-agent System (MAS). MAS-Bench provides pre-built maps and methods so that users can focus on evaluating the methods. MAS-Bench will be a starting point for the evaluation of suitable methods in each field.

Baseline of optimization methods
MAS-Bench is already tried five optimization methods:
1. Random Search (RS)
2. Particle Swarm Optimization (PSO)
3. Covariance Matrix Adaptation Evolution Strategy (CMAES)
4. Success-History based Adaptive Differential Evolution (SHADE)
5. Tree-structured Parzen Estimator (TPE)

Benchmark problems
We considered four types of benchmark problems:
$2D$, $5D$, $8D$, and $17D$ ("D" is number of parameters).
The optimization becomes harder as D increases.

We are waiting for a method that is better than the five methods.



	2D		5D	
	Observable	Unobservable	Observable	Unobservable
RS	10.37 (1.53)	**0.64 (0.12)**	13.97 (3.48)	10.77 (0.98)
PSO	**6.69 (0.48)**	0.65 (0.21)	8.03 (2.35)	13.34 (5.51)
CMAES	**6.71 (0.53)**	0.80 (0.40)	**5.42 (1.00)**	6.96 (3.69)
SHADE	8.81 (1.09)	0.94 (0.34)	6.12 (0.49)	**4.37 (1.21)**
TPE	7.44 (0.61)	**0.61 (0.16)**	**4.90 (0.13)**	**4.35 (1.60)**

	8D		17D	
	Observable	Unobservable	Observable	Unobservable
RS	17.53 (2.39)	17.37 (0.22)	14.95 (1.85)	16.66 (0.99)
PSO	16.01 (7.23)	16.43 (3.80)	11.06 (0.66)	14.80 (2.01)
CMAES	**5.57 (0.43)**	**9.04 (1.90)**	**9.55 (0.53)**	**13.03 (0.77)**
SHADE	**7.35 (1.15)**	**8.29 (2.80)**	12.02 (1.51)	**11.52 (1.06)**
TPE	7.96 (0.88)	13.38 (3.52)	**9.74 (0.79)**	13.40 (3.16)

Optimization : results of use case

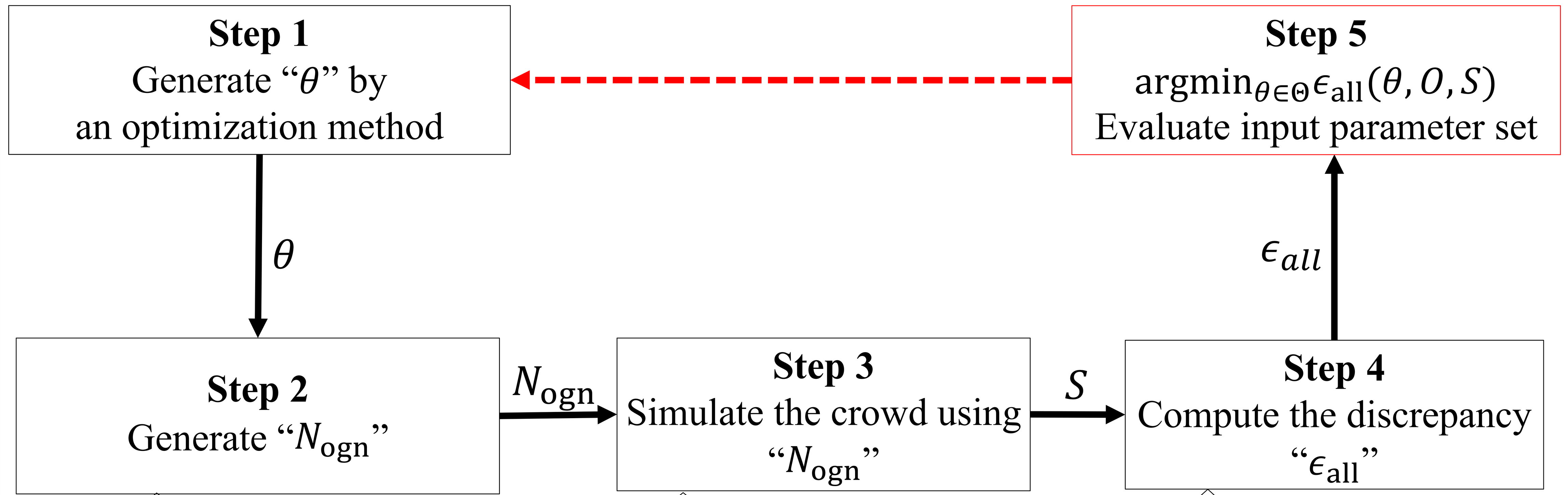
What is MAS-Bench?

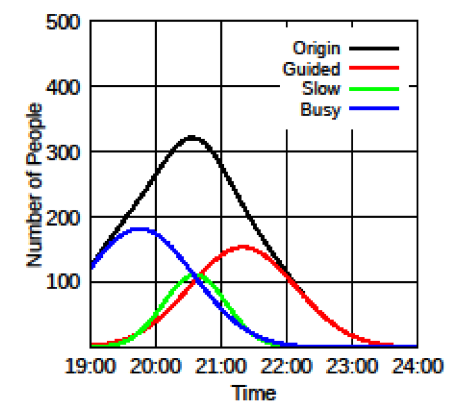
MAS-Bench provides a scenario and modeling of real-world crowd control to exit from a large-scale outdoor festival site. The benchmark handles the following problem to minimize the error between real world observations and its simulated value by adjusting input parameter values. Here, we consider that frequency distribution and walking distance data are given as a ground truth.

$$\theta^* = \mathrm{argmin}_{\theta\in\Theta}\epsilon_{\mathrm{all}}(\theta, O, S)$$

Procedure
Step 1 Generate "θ".
Step 2 Generate "N_{ogn}" from "θ".
Step 3 Collect "S" from "N_{ogn}".
Step 4 Evaluate "ϵ_{all}" from "S" and "O".
Step 5 Evaluate input parameter set.
(Repeat Step1-5.)

Description
$\theta \in \mathbb{R}^D$: A parameter set for gaussian mixture distribution.
N_{ogn} : Origin histogram (plotting origin frequency distribution).
O : The original data (ground truth).
S : The simulation data.
ϵ_{all} : Discrepancy between the original data and simulation data measured by root means square error (RMSE)

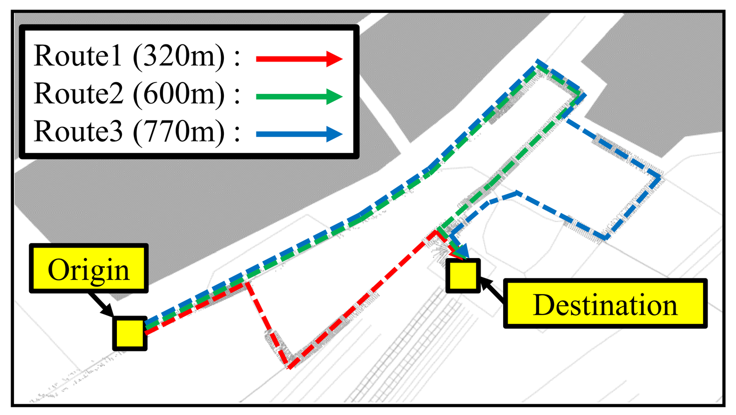




Generated origin histogram of four agent types.

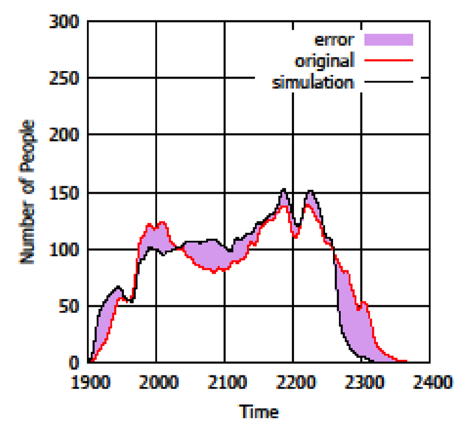
Four agent types

Type	Speed	Route
Busy	Faster than baseline	1
Slow	Slower than baseline	2,3
Guided	Baseline	1,2,3
Pacing	Surroundings mean	1,2,3



Collect simulation data from crowd simulation.

Simulation data
・Destination frequency distribution.
・Walking distance data.



Evaluate discrepancy of input parameter set in simulation and original data.

Original data
・Destination frequency distribution.
・Walking distance data.
・Origin histogram.

Results of Use Case

The results for each baseline are the followings:
1. PSO is strong in the $2D$ problem.
2. TPE is strong in the $5D$ and $17D$ problems.
3. CMAES is strong in the all problems.
4. SHADE ensures higher estimation accuracy for all agent types although it is difficult to estimate the performance of Guided agents because of their frequent route changes.

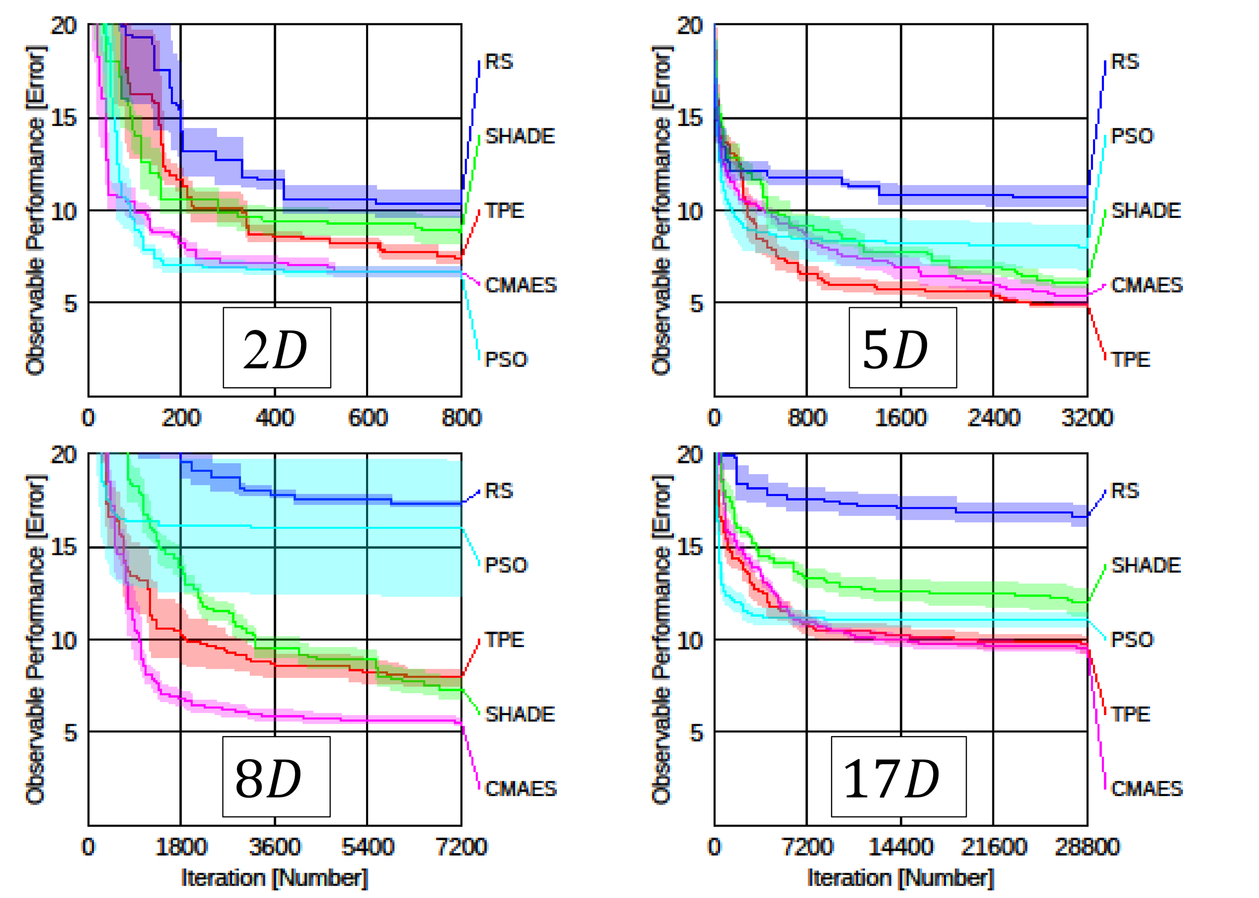


Figure 1 Result of "ϵ_{all}" (observable data).

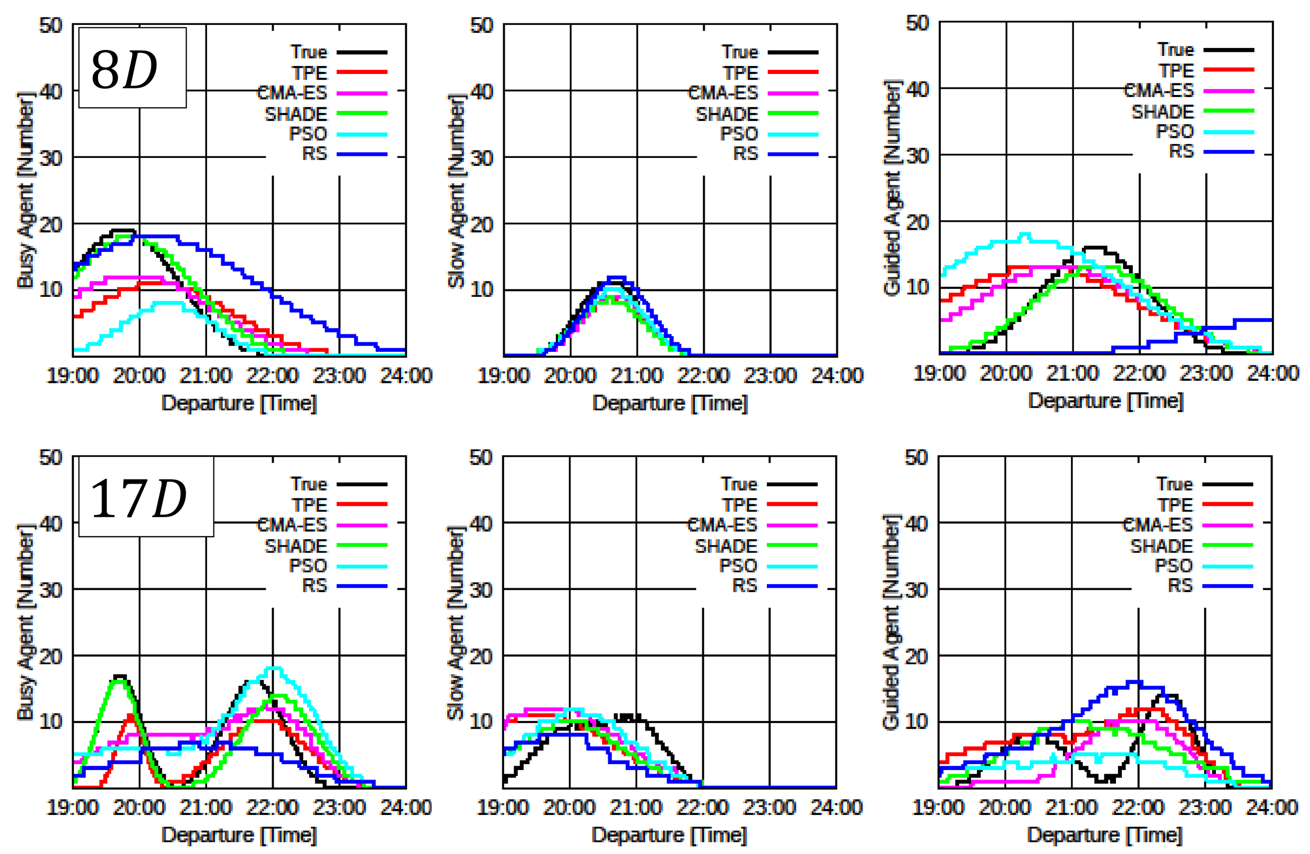


Figure 2 Best performance of origin histogram (unobservable data) of busy (left), slow (center), and guided agent (right) .

**More problems will be added in the future.
Please try your method !**