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

**Extended Abstract**

Shusuke Shigenaka  
University of Tsukuba  
shusuke-shigenaka@aist.go.jp

Shunki Takami  
University of Tsukuba  
s-takami@aist.go.jp

Yuki Tanigaki  
AIST  
tanigaki.yuki@aist.go.jp

Yoshihiko Ozaki  
AIST & GREE, Inc.  
ozaki-y@aist.go.jp

Masaki Onishi  
AIST & GREE, Inc.  
onishi-masaki@aist.go.jp

**KEYWORDS**

Evaluation Tool, Pedestrian Simulator, Parameter Estimation

**ACM Reference Format:**


**1 INTRODUCTION**

Multi-agent crowd simulation is generally used to construct an environment suitable for reality by parameter estimation. This estimation is to modify a model defined by human movements and behaviors to fit real-world data by means of statistics or optimization [2, 11]. In the field of traffic engineering, for example, parameter estimation is used to modify speed models and collision models of pedestrians and vehicles [10, 13]. For different simulation settings, it is not easy to obtain original data of each real-world environment. Few existing estimation methods have been fairly evaluated. Therefore, there is a need for a benchmark to discuss the applicability of certain estimation methods to other use cases.

We propose a benchmark based on a multi-agent system (MAS), MAS-Bench, for parameter estimation on a crowd simulation. The benchmark allows its users to challenge parameter optimization of a complicated real-world system easily. As shown in Figure 1, MAS-Bench provides an integrated benchmark to optimize parameters without any initial construction of simulation models or related systems. We hope that MAS-Bench will boost the further improvement of potential optimization methods, mostly applicable to high-dimensional problems and data assimilation methods required in real-time simulation.

MAS-Bench will be publicly available on GitHub.

**2 PROPOSED BENCHMARK: MAS-BENCH**

**2.1 Problem Setting of Parameter Estimation**

MAS-Bench offers modeling of a real-world crowd routing to exit from a large-scale outdoor festival site. The benchmark handles a problem to minimize the error between an original value $O$ and its simulated value $S$ by adjusting input parameter values. $O$ and $S$ are observable values with pedestrian walking speed from the main festival venue (origin) to the nearest station (destination) and the number of people at the destination. $\theta$ is an unobservable parameter set of the Gaussian mixture model (GMM) to estimate the number of people at the origin. The problem is formulated as follows:

$$\theta^* = \arg\min_{\theta \in \Theta} \epsilon_{\text{all}}(\theta; O, S) \tag{1}$$

The error $\epsilon_{\text{all}}$ is computed as the sum of the linear combination of $\epsilon_{\text{dest}}$ and $\epsilon_{\text{pac}}$, where $\epsilon_{\text{dest}}$ is the error in the number of arrivals, and $\epsilon_{\text{pac}}$ is the error in the walking speed. Each error is calculated as a root means square error (RMSE).

MAS-Bench evaluates an input set $\theta$ of Gaussian distribution parameters: a ratio $\pi$, mean $\mu$, and variance $\sigma$ of the d-th Gaussian distribution. The Gaussian model $f(d; \theta_d)$ of the d-th basis of $f(t; \theta)$ can be described as follows:

$$f_d(t; \theta_d) = \frac{C_d \pi_d}{\sqrt{2\pi\sigma_d^2}} \exp\left(-\frac{(t - \mu_d)^2}{2\sigma_d^2}\right) \tag{2}$$

$$f(t; \theta) = \sum_{d=1}^{D} f_d(t; \theta_d) \tag{3}$$

where $C_d = 1/\int_0^T f_d(t; \mu_d, \sigma_d) dt$ denotes the normalization constant, and $T$ denotes the closing time of the simulation.
Table 1: Performance evaluation of observable and unobservable data using RMSE (mean and variance values after five iterations).

<table>
<thead>
<tr>
<th></th>
<th>2D</th>
<th></th>
<th>5D</th>
<th></th>
<th>8D</th>
<th></th>
<th>17D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Unob</td>
<td>Obs</td>
<td>Unob</td>
<td>Obs</td>
<td>Unob</td>
<td>Obs</td>
<td>Unob</td>
</tr>
<tr>
<td>RS</td>
<td>10.37 (1.53)</td>
<td>0.64 (0.12)</td>
<td>13.97 (3.48)</td>
<td>10.77 (0.98)</td>
<td>17.53 (2.39)</td>
<td>17.37 (0.22)</td>
<td>14.95 (1.85)</td>
<td>16.66 (0.99)</td>
</tr>
<tr>
<td>PSO</td>
<td>6.69 (0.48)</td>
<td>0.65 (0.21)</td>
<td>8.03 (2.35)</td>
<td>13.34 (5.51)</td>
<td>16.61 (7.23)</td>
<td>16.43 (3.80)</td>
<td>11.06 (0.66)</td>
<td>14.80 (2.01)</td>
</tr>
<tr>
<td>CMAES</td>
<td>6.71 (0.53)</td>
<td>0.80 (0.40)</td>
<td>5.42 (1.00)</td>
<td>6.96 (3.69)</td>
<td>5.37 (0.43)</td>
<td>9.04 (1.90)</td>
<td>9.35 (0.33)</td>
<td>13.03 (0.77)</td>
</tr>
<tr>
<td>SHADE</td>
<td>8.81 (1.09)</td>
<td>0.94 (0.34)</td>
<td>6.12 (0.49)</td>
<td>4.37 (1.21)</td>
<td>7.35 (1.15)</td>
<td>8.29 (2.80)</td>
<td>12.02 (1.51)</td>
<td>11.52 (1.06)</td>
</tr>
<tr>
<td>TPE</td>
<td>7.44 (0.61)</td>
<td>0.61 (0.16)</td>
<td>4.90 (0.13)</td>
<td>4.35 (1.60)</td>
<td>7.96 (0.88)</td>
<td>13.38 (3.52)</td>
<td>9.74 (0.79)</td>
<td>13.40 (3.16)</td>
</tr>
</tbody>
</table>

2.2 Crowd Simulation

In MAS-Bench, the multi-agent crowd simulator uses actual measurements such as agents’ walking flow data with RGB-D cameras and each walking distance between the origin and destination with a handy Global Positioning System (GPS). Both were obtained every 15 minutes on Route 1 (320m), Route 2 (600m), and Route 3 (770m). These different lengths of routes are provided to control the flow on each route and to alleviate congestion. The crowd simulator sets four types of walking agents according to walking speed and roles: Guided Agent as the speed baseline, Busy Agent at 10% faster than the Guided Agent, Slow Agent at 50% slower, and Pacing agent at an average walking speed within one link (about 30m).

2.3 Baseline of Optimization Methods

MAS-Bench provides five optimization methods: Random search (RS), Particle swarm optimization (PSO), Covariance Matrix Adaptation Evolution Strategy (CMA-ES), Success-History based Adaptive Differential Evolution (SHADE), and Tree-structured Parzen Estimator (TPE).

RS is an optimized baseline algorithm to randomly generate an input parameter set, specifically for a high-dimensional and multimodal function [3]. PSO is a typical metaheuristic algorithm to seek an optimal solution according to adaptability to a given environment [6]. CMA-ES is a state-of-the-art method of black-box optimization [7–9] to generate an input parameter set from a multivariate normal distribution of a suitable parameter set. SHADE is also known as a state-of-the-art method of black-box optimization [12], derived from Joint Approximation Diagonalization of Eigen-matrices (JADE) [14]. TPE is a high-performance algorithm adopted for Bayesian optimization by Hyperopt [4] and Optuna [1] to probabilistically determine the next input parameter set from the evaluation history [5].

3 EXPERIMENTS: USE CASE OF MAS-BENCH

This section demonstrates a use case of MAS-Bench parameter optimization with the experimental results.

3.1 Experiment Settings

To compare the performance of optimization methods, we evaluated \((D + 1)/3^2 \times 20\) input parameter sets \(x \times 40\) generations where \(D\) is the number of dimension (parameters). Each dimension is set based on calculation of [distribution]*[agent type]*[distribution parameter set]. We considered four types of benchmark problems.

- 2D for Guided Agent
- 5D for Guided Agent and Busy Agent
- 8D for one distribution of three agent types
- 17D for two distributions in one agent type

The total number of walking agents is 45,000 for a large-scale crowd problem and 4,500 for a small-scale crowd problem. The number of pacing agents is set to 900 for a large-scale and 90 for a small-scale. The departure agent ratio is 0.01 for both problems. We used a calculator of Intel CPU Core i9-9900K (3.5GHz, 10 cores, 20 threads), which can calculate up to 20 input parameter sets simultaneously per calculation or one minute for an input parameter set.

3.2 Results of MAS-Bench Use Case

Table 1 lists the results of evaluating the small-scale crowding problem (Number of the walking agents 4,500) by dimension, showing the mean and variance values after five iterations of optimization. The best solution among the five optimization methods is shown in red bold, and the next best solution is shown in bold. From the experimental results, for observable data, CMA-ES performed better than the other algorithms. For unobservable data, TPE is better for low-dimensional problems (2D and 5D), and SHADE is better for high-dimensional problems (8D and 17D). For RS and PSO, the results for Observable and Unobservable are both poor except for 2D. This suggests that this benchmark can estimate parameters closer to the original values by minimizing Observable data’s error.

4 CONCLUSION

This study proposed MAS-Bench, a parameter optimization benchmark to enable users to evaluate their optimization methods. MAS-Bench provides MAS-simulation modeling of crowd flows from the origin to the destination of the small-scale and large-scale problems. The MAS-Bench crowd simulator controls the crowd flows by varying in a set of input parameters. Users can optimize the unobservable parameter set by applying their optimization methods to MAS-Bench to fairly analyze the candidate methods under the same model and settings.

The present MAS-Bench uses estimates by the simulator as original values. In the future study, we will introduce MAS-Bench using original values and examine the multidisciplinary correlation of real-world phenomena.
REFERENCES


