R13/Sarthak Mittal/200050129

March 26, 2023

This paper simulates **Curling** game, and focuses on finding the **optimal hammer shot**, the deciding action before the score tally. The setting is a continuous low-dimensional space with stochastic outcomes. The authors formulate a method called **Delaunay Sampling (DS)** which is more **efficient** and outperformed existing algorithms as well as **Olympic** performance.

The last shot (called "hammer shot") of a round (called "end") of Curling heavily influences the score of the end. Though this shot does not need reasoning about the opponent, the state and action spaces are **continuous**, the action outcomes are **stochastic**, and the score function is **non-convex**. The action space can be modelled using the angle θ with the x-axis and velocity v with the y-axis. Due to the constraint of execution being stochastic, the **expected value** of a shot can only be computed using conditional weights.

Among related works, the authors explored Hierarchical Optimistic Optimization (HOO) in the context of Continuous Bandits, Gaussian Process Optimization (GPO), Particle Swarm Optimization (PSO) and Covariance Matrix Adaptation - Evolution Strategy (CMA-ES). The works related to Billiards (a similar game with continuous state and action spaces, and stochastic outcomes) used **domain knowledge** to create discrete actions.

To tackle the non-convex optimization and continuous space, the authors extend the **Delaunay Triangulation (DT)** to **discretize** the action space followed by sampling in **"promising"** regions. The promising regions are treated as **"arms"** in a multi-armed bandit problem. The authors also developed a **simulator** to evaluate the selection based on DS and other representative algorithms.

The **state** of the shot is determined by: (1) the **score** differential, (2) the number of **ends** left, (3) the **positions** of rocks in play. The rock is thrown at some angle θ relative to the centre line with an initial linear (v) and angular (ω) velocity, causing the rock to trace an **arc** whose direction is determined by the sign of ω. Two players move along with the rock; one can choose to **sweep** the rock at any time. The scores are added after all rocks come to **rest** depending on the number of rocks that are **closer** to the "house centre" than any rock of the opposition. The reasons why the intended outcome may not be achieved are: (1) **human error** (thrower may not choose perfect parameters, sweeps may be misapplied, etc.), (2) **field variations** (the ice and rocks are not uniform).

The authors simplify the action space by only considering the **direction** of angular velocity and accounting for sweeping into the simulator **execution** model. The physical simulation is based on **Chipmunk 2D rigid body physics** library. The execution model represents the **variability** in outcomes as a **stochastic transformation**. The primary purpose of **sweeping** is the reduction of execution error which is implicitly modelled as an increase in the **likelihood** of the trajectories matching. The intended shot parameters are **perturbed** using samples from a heavy-tailed, zero-mean, **student-t** distribution tuned to Olympic-level.

The algorithm comprises a **sampling stage** and a **selection stage**. In the sampling stage: (1) sample the stochastic **objective** at uniformly distributed points, (2) apply DT for **partitioning**, (3) assign **weights**, (4) sample regions with weights as **probabilities**, (5) sample a point from the region along with its **value** of objective, (6) repeat from step 2 for T **iterations**. The weights are defined using the **area** and the **score** of the region such that more weight is put on **higher-valued** regions as t increases. In the selection stage: (1) assign new **weights** based on rewards, (2) select the k**max-weight** regions, (3) run $\hat{T}$ iterations of **UCB** using the regions as arms and sample the **objective** at the incentre (4) return the incentre of the **most sampled** region.

The first stage is **exploratory** and thus has an optimistic score, while the second stage is to **exploit** the most significant expected value. When combined, DT and UCB seem to cover each other's **weaknesses**. The authors focus on optimizing the **win percentage (WP)** instead of expected points (EP). The state is modelled by n (the number of ends left to play) and δ (the lead of the hammer shot team). The experiments were based on the data from **2010 Winter Olympics**. With slight variations, **simulations** were performed for DS, HOO + UCB, PSO, BIPO-CMA-ES and GPO. The **Wilcoxon Signed-Rank Test (WSRT)** confirmed that the performance of DS was not by chance.

Comparing the win percentages showed that DS **outperformed** all the baselines and Olympic teams. Detailed analysis showed that there were instances where DS chose a shot which was far more **superior** and promising than the Olympic teams. The authors pointed out a few **limitations** of this comparison: (1) the **calibration** to Olympic level may not be entirely accurate, (2) the simulator does not perfectly replicate the **physical conditions**, (3) data from Olympics was logged **manually**. These factors could work in favour of DS or might adversely affect it too.