Preprints of Control Conference Africa 2024
Balaclava, Mauritius, 16-17 September 2024

Grinding Mill Optimisation using Grind Curves and Reinforcement Learning

Jacques Olivier*. William J. Shipman*

**Measurement and Control Division, Mintek, Johannesburg, South Africa*
(email: jacqueso@mintek.co.za ; williams@mintek.co.za)

Abstract: This work presents a real-time optimiser for grinding mills based on grind curve relationships. The optimiser is a continuum-armed bandit algorithm from the bandit class of problems in reinforcement learning. The optimiser can manage complex objective functions with multiple constraints and time-varying components. The optimiser uses the uncertainty in the underlying models to guide exploration. Optimiser testing occurred in a simulation environment containing a dynamic grinding-mill model. The optimiser demonstrated the ability to steer the mill throughput near theoretical optimums defined by grind curves.

Keywords: real-time optimisation; grinding mill; reinforcement learning; process control; grind curve

1. INTRODUCTION

Grinding mills are responsible for the size reduction of ores in mineral-processing plants. The process involves rotating large-diameter drums with large masses of ore and steel grinding media, making grinding the largest energy consumer in processing plants (Napier-Munn, 2015). Grinding mill operation is highly sensitive to fluctuations in ores' competency and particle size, causing frequent disturbances and suboptimal operation.

Practitioners generally agree that grinding mills follow grind curve behaviour at steady-state (Powell et al., 2009). Grind curves relate mill-performance indicators, such as the mill power, throughput and discharge particle size (grind), to the mill load and speed. Grind curves are generally unknown during regular operation due to the uncertainty in ore feedstocks and the state of the internal mill liners. However, we can employ peak-seeking algorithms to reconstruct these curves from process data by exploiting our understanding that grind curves exhibit a quadratic shape. Previous work has demonstrated this using extremum-seeking control (ESC) (Ziolkowski et al., 2022).

This work explores using reinforcement learning (RL) to optimise grinding mills through direct interaction with the process. Continuum-armed bandit (CAB) agents were trained to recover the unknown grind curves (Sutton and Barto, 2018, chap. 2). The CABs function as real-time optimisers, providing setpoints to regulatory controllers. The optimiser was tested in a simulation environment containing a non-linear, dynamic grinding-mill model calibrated to grind curve data (Le Roux et al., 2020).

2. METHODOLOGY

In a typical RL workflow, an agent takes an action $\boldsymbol{A}_t$ at every timestep t. An environment, in this case the dynamic grinding-mill model, responds to the action and sends a reward $\boldsymbol{R}_t$ back to the agent. $\boldsymbol{R}_t$ can be a single scalar reward, or a vector of rewards. The agent learns to maximise the expected reward by modelling the reward as a function of the actions.

The CAB agent models the reward vector as an arbitrary combination of real-value and binary components. Real-value components, such as mill-performance indicators, are modelled as polynomials of known order using robust linear regression. Binary components indicate constraint violations in the environment and are used to train logistic regression models. The decision boundaries spanned by these models are used as inequality constraints by the optimisation algorithm for action selection.

The CAB agent uses robust, rolling linear regression to model the polynomial relationship between mill-performance indicators and actions taken. The input feature vector is given in equation (1), with ω indicating the polynomial expansion of the variables in the action space. The action space $\boldsymbol{a}_t$ consists of the mill filling setpoint, $J_{T,SP}$, and speed setpoint, $u_{\phi,SP}$.

$$\omega(\boldsymbol{a_t}) = \{J_{T,SP}^2, u_{\phi,SP}^2, J_{T,SP}u_{\phi,SP}, J_{T,SP}, u_{\phi,SP}, 1\} \quad (1)$$

Regression models were trained using an iteratively reweighted least squares algorithm with Huber's t-function (Huber, 1964) as robust criterion estimator. A separate model of the form in equation (2) is trained for each of the $N_{\mathbb{R}}$ real-value reward components in $\boldsymbol{R}_t$.

$$R_{t,j} = \boldsymbol{\theta}_j^T \omega(\boldsymbol{a}_t); \quad j \in \mathbb{Z}^{1 \le j \le N_{\mathbb{R}}} \quad (2)$$

Thompson sampling (Thompson, 1933) was used to incorporate uncertainty into the agent's decision-making. The process involves generating a Gaussian distribution of potential models using the trained model parameters and covariance, according to equation (3).

$$p(\boldsymbol{\theta}_j) = \mathcal{N}(\boldsymbol{\theta}_j, \mathrm{cov}(\boldsymbol{\theta}_j)) \quad (3)$$

At each step, a candidate parameter vector $\widehat{\boldsymbol{\theta}}_j$ is sampled from $p(\boldsymbol{\theta}_j)$ to build the polynomial model representing each grind

curve. Intuitively, Thompson sampling uses the uncertainty in the trained model parameters to explore the action space.

The environment detects constraint violations and communicates the outcome via binary components in the reward vector. Once a constraint is violated, a logistic regression model of the form in equation (4) is trained. The left-hand side of equation (4) evaluates to the distance of the new sample to the decision boundary defining the constraint.

$$g_j(\boldsymbol{a}_t) = \boldsymbol{w}_j^T \boldsymbol{a}_t + b_j \tag{4}$$

The individual real-value component models are aggregated into a scalar reward function $\mathcal{R}_t$ according to the user-specified objective function $\lambda(\boldsymbol{\mathcal{L}})$. $\boldsymbol{\mathcal{L}}$ represents the set of trained regression models, now each parametrised by $\widehat{\boldsymbol{\theta}}_j$. Action selection involves solving the optimisation problem defined below. $\boldsymbol{a_u}$ and $\boldsymbol{a_l}$, represent the upper and lower limits on each of the variables in $\boldsymbol{a}_t$.

$$\max_{\boldsymbol{a_t}} \quad \mathcal{R}_t \tag{5}$$

$$\text{subject to} \quad \boldsymbol{w}_j^T \boldsymbol{a_t} + b_j \geq 0; j \in \mathbb{Z}^{1<j\leq N_b} \tag{6}$$

$$\boldsymbol{a_l} \leq \boldsymbol{a_t} \leq \boldsymbol{a_u} \tag{7}$$

$$|\boldsymbol{a_t} - \boldsymbol{a_{t-1}}| \leq \delta(\boldsymbol{a_u} - \boldsymbol{a_l}) \tag{8}$$

The CAB algorithm is parameterised by the step size and step frequency, represented by δ and N_{sf}, respectively. δ represents the maximum allowable change in the actions between successive steps, while N_{sf} represents the number of steps made by the agent, per step made by the environment. N_{sf} is required for the process to reach or approach steady-state before the agent makes another step.

3. RESULTS

The agent was tested in a simulation environment identical to the one described in Ziolkowski (2022). The environment is a simulation of a typical grinding mill with regulatory control on the mill filling and discharge density using proportional-integral (PI) controllers. Artificial noise was added to the throughput, grind and filling signals.

The CAB agent was configured to maximise the mill throughput, while adhering to constraints on the grind size achieved. The environment was configured to produce the reward vector in equation (9), containing only the throughput as a real component and the two binary components indicating upper and lower limits on the grind.

$$\boldsymbol{R}_t = \langle throughput_t, bin_{grind<0.5}, bin_{grind>0.93} \rangle \tag{9}$$

The trajectory of the agent through the simulation is depicted in Figure 1. This agent used a hyper-parameter set of $\langle \delta, N_{sf} \rangle$ equal to $\langle 0.07, 48 \rangle$, which was identified as optimal through a grid search. However, large regions in the hyper-parameter space yielded similar results, indicating a degree of robustness to the hyper-parameter settings.

Figure 1 demonstrates how the agent learns to maximise the mill throughput by increasing the speed to 0.75 and decreasing the filling to around 0.36. This is near the (speed, filling) configuration of (0.75, 0.35), where the theoretical maximum throughput of the model is achieved (Ziolkowski et al., 2022, tbl. 2), thereby proving that the agent is capable of finding regions near the unknown optimums.

A comparison with the results of Ziolkowski et al. (2022) revealed that the CAB agent reaches similar steady-state values to ESCs in comparable timeframes. While based on different fundamentals, there is practically a high degree of alignment between the two methods. The main difference is the stochastic exploration mechanism of the CAB versus the deterministic, periodic perturbations generally used in ESC. Incorporating the model uncertainty into the exploration step should increase the optimiser's robustness to noisy and faulty data often encountered in process systems.

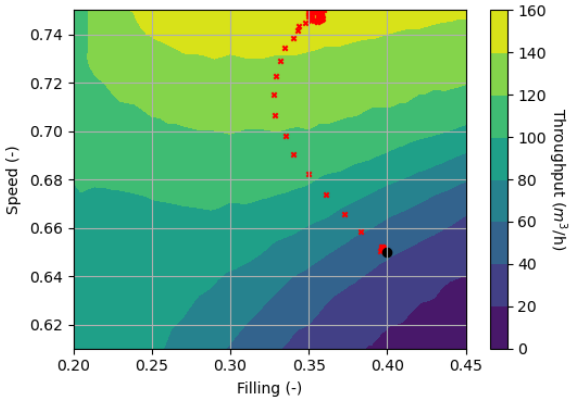

Figure 1: Average trajectory of the optimal CAB agent on the throughput grind curve for throughput optimisation. The black marker indicates the agent's initial position.

REFERENCES

Huber, P.J., 1964. Robust Estimation of a Location Parameter. Ann. Math. Stat. 35, 73–101.

Le Roux, J.D., Steinboeck, A., Kugi, A., Craig, I.K., 2020. Steady-state and dynamic simulation of a grinding mill using grind curves. Miner. Eng. 152, 106208. doi:10.1016/j.mineng.2020.106208

Napier-Munn, T., 2015. Is progress in energy-efficient comminution doomed? Miner. Eng. 73, 1–6. doi:10.1016/j.mineng.2014.06.009

Powell, M.S., Van der Westhuizen, A.P., Mainza, A.N., 2009. Applying grindcurves to mill operation and optimisation. Miner. Eng. 22, 625–632. doi:10.1016/j.mineng.2009.01.008

Sutton, R.S., Barto, A., 2018. Reinforcement learning: an introduction, Second edition. ed, Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts, London, England.

Thompson, W.R., 1933. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. Biometrika 25, 285–294. doi:https://doi.org/10.1093/biomet/25.3-4.285

Ziolkowski, L., Le Roux, J.D., Craig, I.K., 2022. Extremum seeking control for optimization of an open-loop grinding mill using grind curves. J. Process Control 114, 54–70. doi:10.1016/j.jprocont.2022.04.005