

# Partially Linear Contextual Bandits

Minsu LEE<sup>1</sup> and Sunyoung SHIN<sup>2</sup>

1) *Department of Mathematics, Pohang University of Science and Technology, Pohang 37673, Korea, crisophy@postech.ac.kr*

2) *Department of Mathematics, Pohang University of Science and Technology, Pohang 37673, Korea, sunyoungshin@postech.ac.kr*

Corresponding Author : Sunyoung SHIN, sunyoungshin@postech.ac.kr

## ABSTRACT

A contextual bandit is a sequential decision-making framework in which, at each round, an agent observes contextual information about the available actions, selects one action, and receives a corresponding reward. To choose the optimal action, the agent infers the expected reward for each possible action in advance using the given context. It is crucial to specify an appropriate context–reward relationship in the contextual bandit framework, as the accuracy of decision-making depends on how effectively this underlying relationship is modeled. Traditional methods for modeling the relationship typically employ either fully parametric or non-parametric approaches, each with its own set of trade-offs. Parametric approaches are computationally efficient and simple to implement, however they rely on strict assumptions that often are not met in complex real-world scenarios. Non-parametric methods offer greater flexibility, but may not fully leverage clear underlying patterns that could otherwise improve prediction. In this work, we bridge these two paradigms by proposing a partially linear model for the expected reward. For given contextual features  $c_t = (x_t^T, z_t^T)^T \in \mathbb{R}^{p+q}$  at round  $t$ , the expected reward is modeled by a linear function of  $x_t$  and a nonparametric function of  $z_t$ . We employ kernel regression to estimate the non-parametric component  $f_*(z_t)$ . This approach incorporates prior knowledge about some contextual features having a linear effect on the reward while remaining flexible about potentially complex relationships between the other features and the reward. We introduce a novel algorithm, PartLinUCB, designed for this partially linear model, and provide a theoretical regret bound that matches the regret bounds of KernelUCB and LinUCB when  $p = 0$  and  $q = 0$ , respectively. Empirical studies demonstrate that PartLinUCB achieves superior performance compared to representative baseline methods, validating both the theoretical framework and practical utility of the partially linear approach.