ICML
International Conference On Machine Learning

# *Priority-Aware Shapley Value*

**Kiljae Lee, Ziqi Liu, Weijing Tang, and Yuan Zhang**

The Ohio State University, Carnegie Mellon University

The Ohio State University

Carnegie Mellon University

# Data Shapley: valuing each training point



**Goal.** Quantify how much each data point contributes to model performance.

- Given a dataset $\mathcal{D} = \{z_1, \dots, z_n\}$ and utility $U: 2^{\mathcal{D}} \to \mathbb{R}$ ,Ghorbani and Zou (2019) proposed the **Data Shapley**:

**Definition [Shapley Value (Data Shapley)]**

$$\phi_i(U) = \mathbb{E}_{\pi \sim \mathrm{Unif}(\Pi)}[U(P_i^\pi \cup \{z_i\}) - U(P_i^\pi)],$$

where $P_i^\pi$ is the set of data points that appear before $z_i$ in the random permutation of $\mathcal{D}$, denoted by $\pi \in \Pi$.

- $U(S)$ measures model performance when trained on the subset $S$.

# Classical Shapley Axioms



- The Shapley value is the unique value satisfying **four classical axioms**:

| Efficiency | Linearity | Null Player | Symmetry |
|---|---|---|---|

**Axiom [Symmetry (informal)]**

**same marginal contribution $\Rightarrow$ same value**

Operationally, this leads to a **uniform average over all player orders**:

$$\pi \sim \mathrm{Unif}(\Pi).$$

***But should every player always be treated as exchangeable?***

# Example: Data Augmentation

But data contributors are **often not exchangeable**.

## Data Lineage

**Original Data**

↓

**Augmented Data**

An augmented point should not be evaluated as if it were independent of its source.

## Trust and Risk

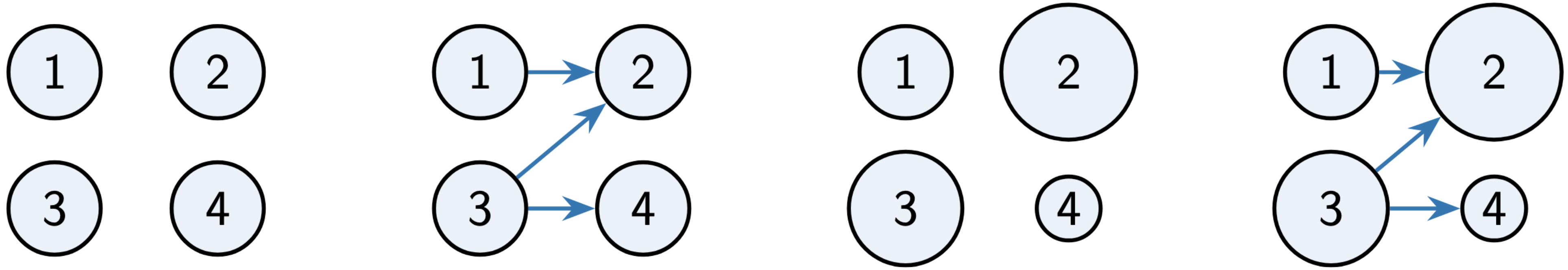
| **Useful Synthetic** | **Pure Copy** | **Poisoned Synthetic** |
|---|---|---|

Even among augmented points, some data may be less trusted, or even harmful.

***We need Shapley attribution reflecting these prior information.***

# Two Kinds of Priority

There are two kinds of priority which **captures these asymmetry**.




No Priority | Hard Priority | Soft Priority | **PASV**

$\pi \sim \mathrm{Unif}(\Pi)$

$\pi \sim \mathrm{Unif}(\Pi^G)$

$\pi \sim \prod_{t=1}^{n} \frac{\lambda_{\pi_t}}{\sum_{s \leq t} \lambda_{\pi_s}}$

$\pi \sim ?$

**Hard Priority:**
Directed Acyclic Graph (DAG; $\mathbf{G}$) rules out invalid orders

**Soft Priority:**
Weights vector ($\boldsymbol{\lambda}$) encodes trust/risk

# Priority-Aware Shapley Value



**Definition [Priority-Aware Shapley Value (PASV)]**

***PASV*** *of $i$-th player is defined as*

$$\phi_i^{\mathrm{PASV}}(U) = \mathbb{E}_{\pi \sim p^{(\mathrm{G},\lambda)}}[U(P_i^\pi \cup \{z_i\}) - U(P_i^\pi)],$$

*where*

$$p^{(\mathrm{G},\lambda)}(\pi) \propto \prod_{t=1}^{n} \frac{\lambda_{\pi_t} \left|\max_{\mathrm{G}}(S_t)\right|}{\sum_{k \in \max_{\mathrm{G}}(S_t)} \lambda_k}, \qquad \pi \in \Pi^G.$$

**Hard Priority:** orders violating DAG **G** have probability zero.

**Soft Priority:** admissible choices are biased by weight $\boldsymbol{\lambda}$.

# PASV is Canonical and Computable

## Boundary cases

- If there is no priority structure, PASV reduces to **Shapley Value**
- If $\lambda_1 = \cdots = \lambda_n$, PASV reduces to **Precedence Shapley Value (Faigle and Kern, 1992)**.
- If the DAG has a special form, PASV reduces to **Weighted Shapley Value (Kalai and Samet, 1987)**.

## Practical estimation

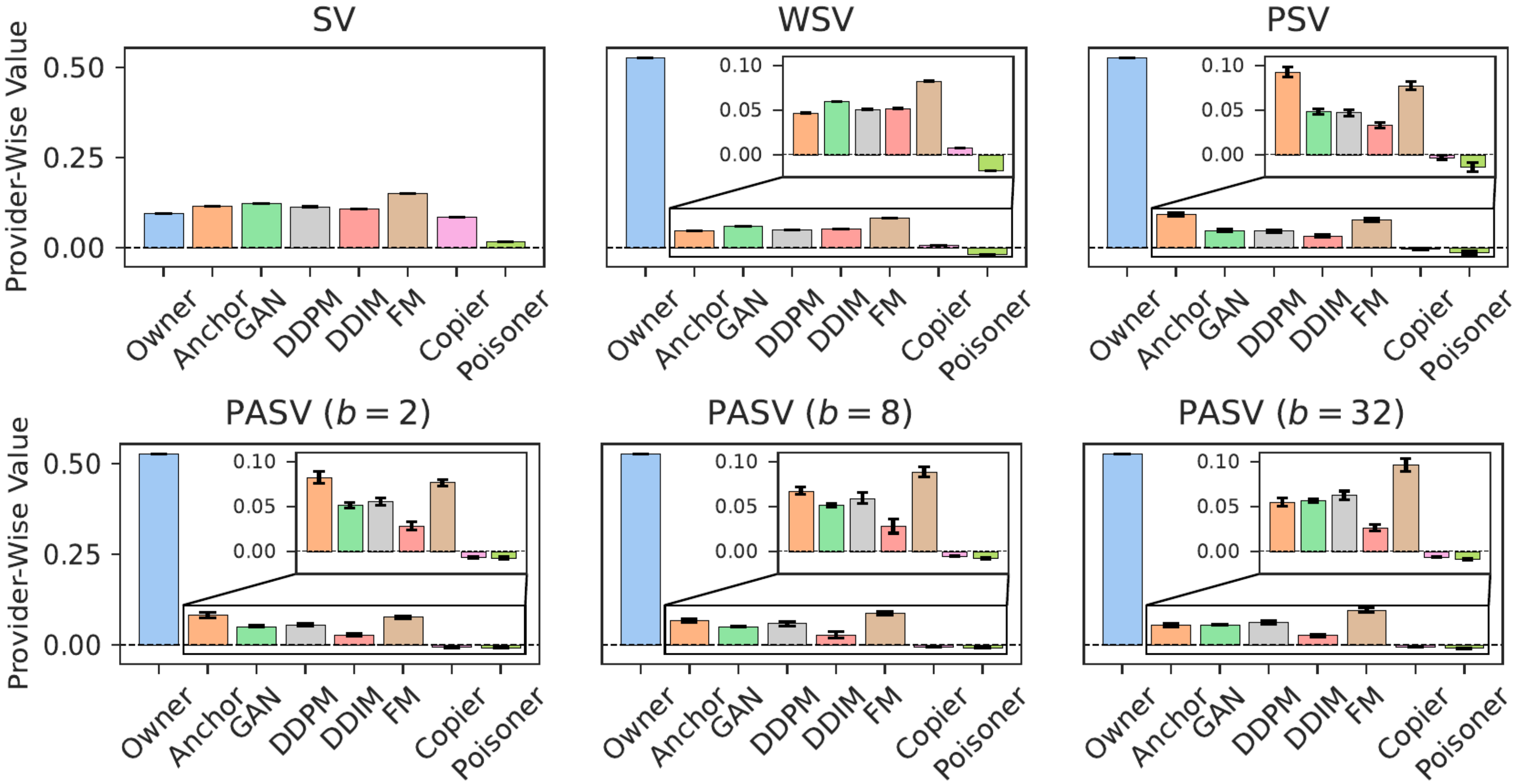
- **Natural axioms** uniquely characterize the PASV order distribution.
- **Adjacent-swap Metropolis–Hastings** samples feasible orders for general DAGs.

# Experiment: Lineage-Aware Data Valuation

**Setup.** Several data providers collaborate to train an image classifier for MNIST.

**Providers.**

- Owner: Original data
- Anchor: Classical augmentation
- Boosters: Augmentation with GenAI
- Copier: Pure copy of data
- Poisoner: Poisoned copy by label-flipping




Priority-Aware Shapley Value

# Takeaway



**Shapley values uniformly average over player orders.**

**PASV makes those orders priority-aware:**

- **Hard priority:** general precedence DAGs rule out invalid orders.
- **Soft priority:** individual weights encode trust, risk, or preference.

***PASV respects lineage and softly adjusts credit by trust/risk.***