Energy Based Learning for Cooperative Games,

with Applications to Valuation Problems in Machine Learning

Yatao Bian*, Yu Rong, Tingyang Xu, Jiaxiang Wu,
Andreas Krause, Junzhou Huang

Tencent AI Lab
ETH Zurich

2022-03

* https://yataobian.com/

Background: valuation problems in ML & player valuations

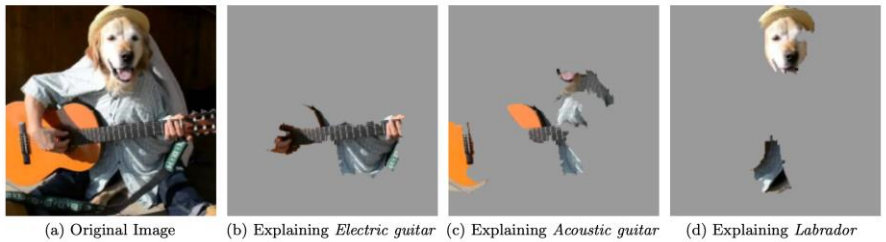
Tencent AI Lab ETH zürich

Valuation problems in ML
- ❑ Feature interpretation
- ❑ Data valuation
- ❑ Model valuation for ensembles

Player valuations in cooperative games:
- ❑ Shapley value
- ❑ Banzhaf value

Cooperative game $(N, F(S))$:
$N = \{1, \dots, n\}$: n players
$F(S)$: payoff of a coalition S

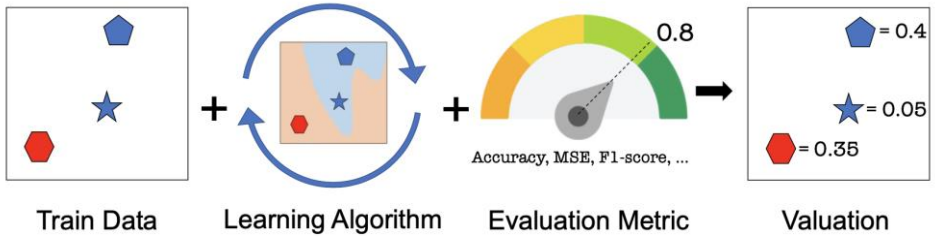
one feature ⟺ one player



(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

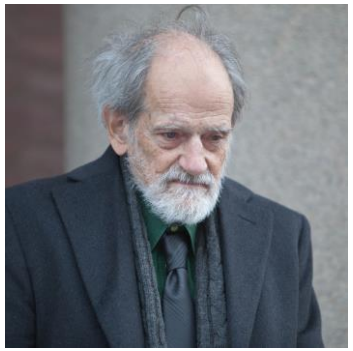
one sample ⟺ one player



Player valuation: assign importance to players

Shapley value [1953] (2012 Nobel Memorial Prize)

$$\mathrm{Sh}_i = \sum_{S \subseteq \mathcal{V} \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} [F(S \cup \{i\}) - F(S)]$$



Lloyd Stowell Shapley
(1923 ~ 2016)

I. Covert, S.Lundberg & S. Lee. "Explaining by removing: A unified framework for model explanations. JMLR 2021.

A. Ghorbani & J. Zou. Data shapley: Equitable valuation of data for machine learning. ICML 2019.

Overview of the Proposed Variational Values

Player valuations in cooperative games

Valuation problems in ML

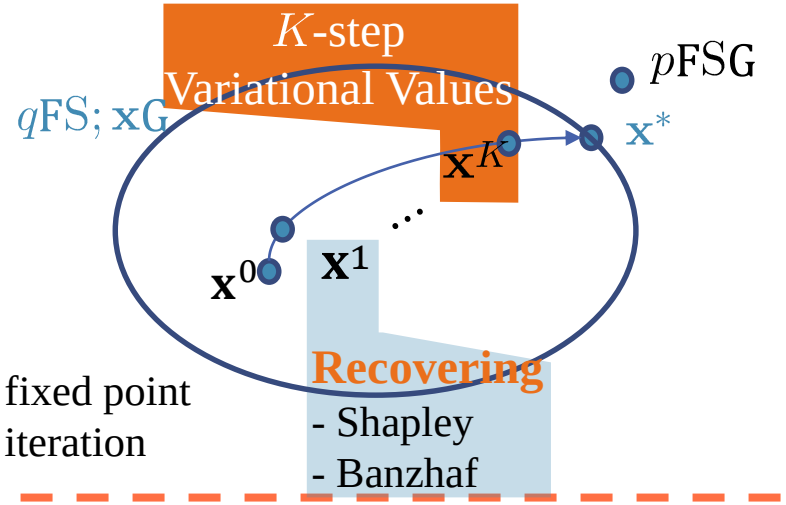
- ❑ Feature interpretation
- ❑ Data valuation
- ❑ Model valuation

Energy Based Model for a cooperative game (N, v_{FSG}):

$p_{\text{FSG}} \propto \exp v_{\text{FSG}}/T_{\text{G}}$

$$\min_{\mathbf{x}} \mathrm{KL}(q(\mathrm{S};\mathbf{x}) \,||\, p(\mathrm{S}))$$

Mean field variational parameters $\mathbf{x}$ as **principled player valuations**



K-step Variational Values

Axioms

- ✓ Null player
- ✓ Marginalism
- ✓ Symmetry

One step → Shapley, Banzhaf
Multi step → Lower decoupling error

Principled valuation criteria

Experimental Results

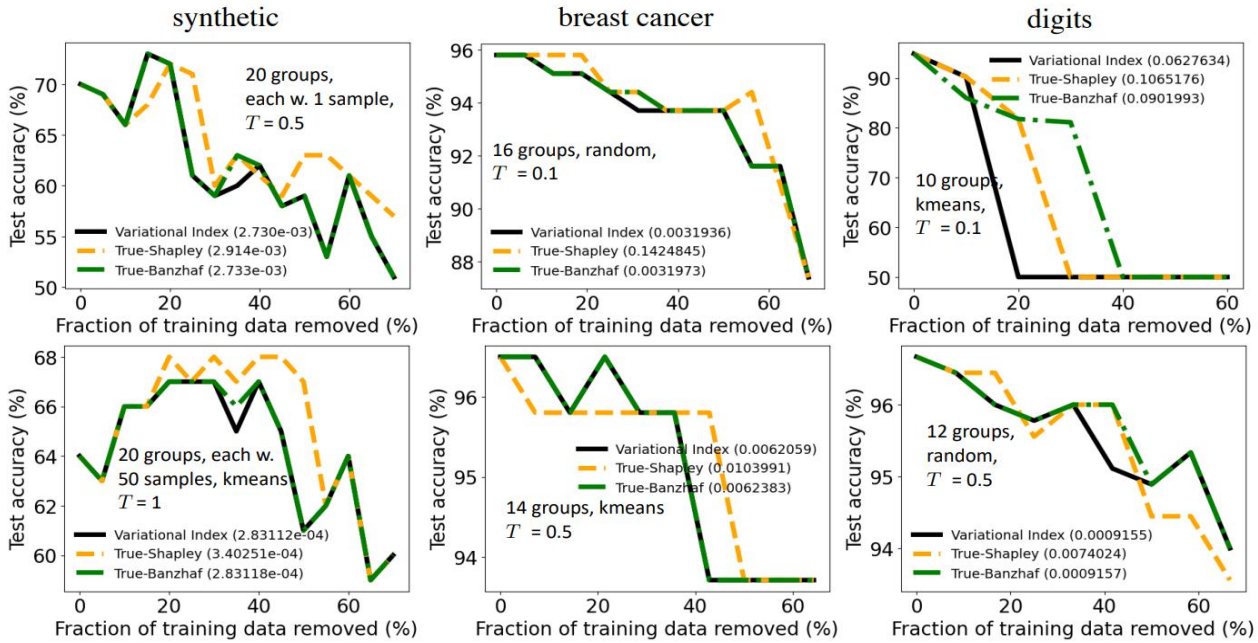
Three groups of experiments:
- ❑ Submodular games
- ❑ Data valuations
- ❑ Feature attributions

Variational Values
achieve lower decoupling error and better valuation performance

★ Code & project page: https://valuationgame.github.io

https://yataobian.com/

data valuation results



feature interpretation results