



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

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 \Leftrightarrow one player

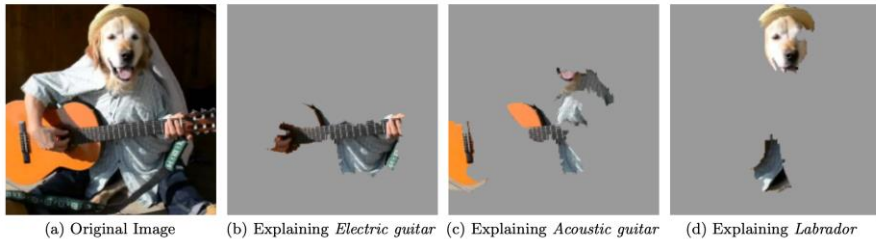
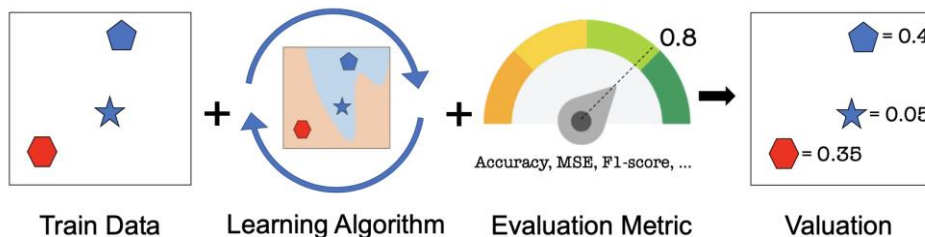


Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

one sample \Leftrightarrow one player



Player valuation: assign importance to players

Shapley value [1953] (2012 Nobel Memorial Prize)

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

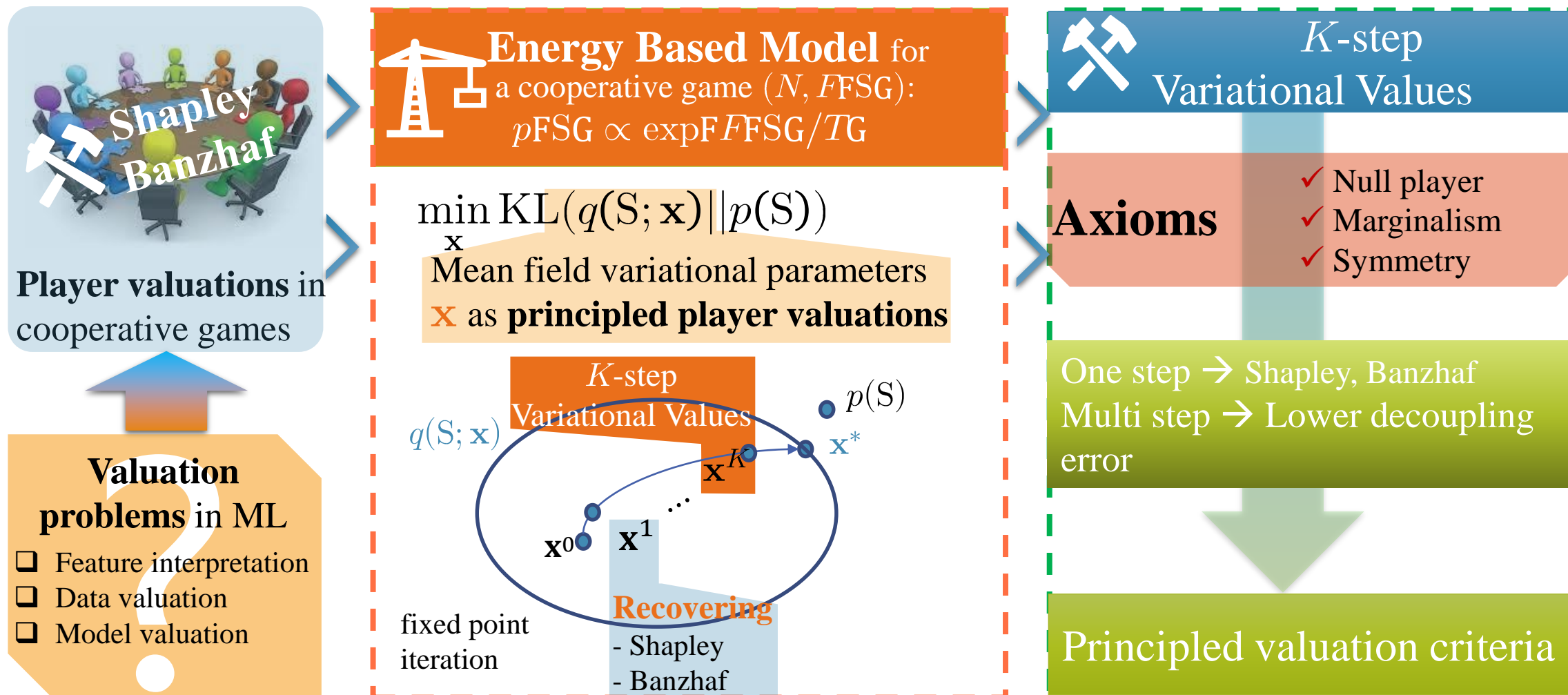


Lloyd Stowell Shapley
(1923 ~ 2016)

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



Experimental Results

Three groups of experiments:

- Submodular games
- Data valuations
- Feature attributions



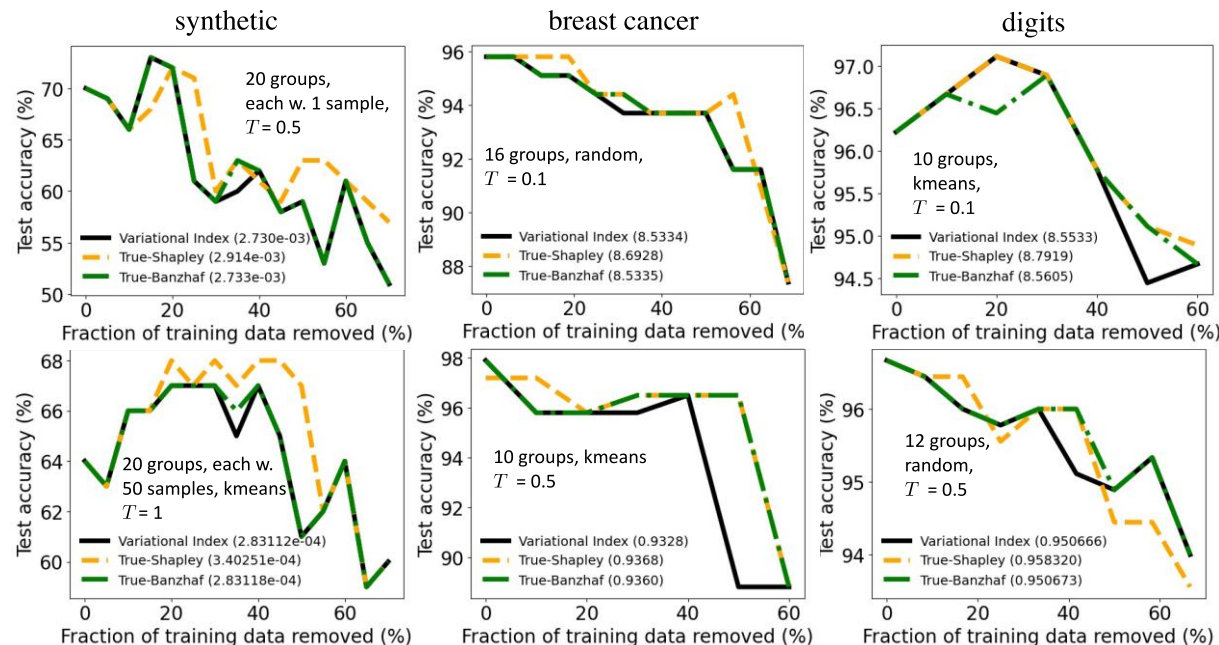
Variational Values
achieve lower decoupling
error and better valuation
performance



Code & project page:
<https://valuationgame.github.io>

<https://yataobian.com/>

data valuation results



feature interpretation results

