

# **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



#### one feature $\Leftrightarrow$ one player



(a) Original Image

ensembles

(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)



Player valuations in cooperative games: □ Shapley value Banzhaf value



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

**Player valuation**: assign importance to players

[1953] (2012 Nobel Shapley value Memorial Prize)

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



Lloyd Stowell Shapley  $(1923 \sim 2016)$ 

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**

#### data valuation results

synthetic



digits

Three groups of experiments:
Submodular games
Data valuations
Feature attributions

96 Variational Index (0.0627634 20 groups, 90 True-Shapley (0.1065176) 8 94 (%) each w. 1 sample, rue-Banzhaf (0.0901993) T = 0.5accuracy ( accuracy 06 16 groups, random, T = 0.110 group kmeans 90 00 ist T = 0.1ariational Index (2.730e-0 Variational Index (0.0031936) True-Shapley (2.914e-03) True-Shapley (0.1424845) True-Banzhaf (2.733e-03) True-Banzhaf (0.0031973 50 50 40 60 40 60 60 20 0 20 0 20 40 Fraction of training data removed (%) Fraction of training data removed (%) Fraction of training data removed (%) **%** 96 8 96 8 66 ) accuracy ( accuracy 6 accuracy 66 12 groups, Index (0.0062059) Variation 20 groups, each w random, True-Shapley (0.0103991) 50 samples, kmean T = 0.5 True-Ban af (0.0062383) T=114 groups, kmeans est T = 0.5iational Index (2.83112 Variational Index (0.0009155) True-Shapley (3.40251e-04) True-Shapley (0.0074024) 94 True-Banzhaf (2.83118e-04) True-Banzhaf (0.0009157 60 40 60 60 20 40 20 20 40 Fraction of training data removed (%) Fraction of training data removed (%) Fraction of training data removed (%) feature interpretation results Variational Shapley Banzhaf xgboost<sup>0.8</sup> Marital Statu ariational Index (0,0055 rue-Shapley (0.0061) - Occupatio e-Banzhaf (0.0061) Hours per wee = 0.81 - Se Data id: 240 – Capital Gai GT label: True 3 other featu 4 Ġ ġ 10 Number of features remove logistic 0.9 18 – Canital Gai Capital Ga - Education-Nu regression0.8 ≥0.7 67 **– Age** = 0.15 O F Hours per week Data id: 7977 - Marital Status GT label: True - Sex Workclass Variational Index (0.0375 True-Shapley (0.0634) 0 - Capital Lo True-Banzhaf (0.1102) 3 other feature ġ 10 6 lumber of features removed

breast cancer

Variational Values achieve lower decoupling error and better valuation performance

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

https://yataobian.com/