

## Review

# The Shapley Value in Data Science: Advances in Computation, Extensions, and Applications

Lei Qin <sup>1,2</sup>, Yingqiu Zhu <sup>1,\*</sup>, Shaonan Liu <sup>1</sup>, Xingjian Zhang <sup>1</sup> and Yining Zhao <sup>3</sup>

<sup>1</sup> School of Statistics, University of International Business and Economics, Beijing 100029, China; qinlei@uibe.edu.cn (L.Q.); 202421412404@uibe.edu.cn (S.L.); 202357044@uibe.edu.cn (X.Z.)

<sup>2</sup> Dong Fureng Institute of Economic and Social Development, Wuhan University, Wuhan 430072, China

<sup>3</sup> Sunwah International Business School, Liaoning University, Shenyang 110036, China; 20226602209@smail.lnu.edu.cn

\* Correspondence: inqzhu@uibe.edu.cn

**Abstract:** The Shapley value is a fundamental concept in data science, providing a principled framework for fair resource allocation, feature importance quantification, and improved interpretability of complex models. Its fundamental theory is based on four axiomatic proper ties, which underpin its widespread application. To address the inherent computational challenges of exact calculation, we discuss model-agnostic approximation techniques, such as Random Order Value, Least Squares Value, and Multilinear Extension Sampling, as well as specialized fast algorithms for linear, tree-based, and deep learning models. Recent extensions, such as Distributional Shapley and Weighted Shapley, have broadened the applications to data valuation, reinforcement learning, feature interaction analysis, and multi-party cooperation. Practical effectiveness has been demonstrated in health care, finance, industry, and the digital economy, with promising future directions for incorporating these techniques into emerging fields, such as data asset pricing and trading.

**Keywords:** Shapley value; data valuation; machine learning interpretability

**MSC:** 91-02; 68-00; 65-00



Academic Editor: Manuel Alberto M. Ferreira

Received: 22 March 2025

Revised: 5 May 2025

Accepted: 9 May 2025

Published: 11 May 2025

**Citation:** Qin, L.; Zhu, Y.; Liu, S.; Zhang, X.; Zhao, Y. The Shapley Value in Data Science: Advances in Computation, Extensions, and Applications. *Mathematics* **2025**, *13*, 1581. <https://doi.org/10.3390/math13101581>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The growing availability of large-scale and high-dimensional data has refocused attention on cooperative game theory techniques, particularly the Shapley value, which is widely used for fair resource allocation, feature importance analysis, and explainable artificial intelligence (XAI). The Shapley value, which was first introduced by Lloyd Shapley (1953) as a solution concept in cooperative game theory [1], provides a principled method for distributing total value among players based on individual contributions. Over the years, it has found applications in a variety of fields, including machine learning [2], economics [3], and network analysis [4].

The Shapley value is a fundamental concept in cooperative game theory, first introduced by [1] and axiomatized by [5], providing a fair allocation method for distributing a total payoff among players based on their contributions. In a cooperative game with  $N = \{1, 2, \dots, n\}$  players and a characteristic function  $v(S)$  assigning a value to each coalition  $S \subseteq N$ , the Shapley value creates a unique and equitable division of the total

surplus. It is defined as the weighted sum of a player's marginal contributions across all potential coalitions:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} [v(S \cup \{i\}) - v(S)], \quad (1)$$

where  $v(S \cup \{i\}) - v(S)$  represents the marginal contribution of player  $i$  to coalition  $S$ , and the weight ensures that all orderings of players are considered equally. This formulation reflects the idea that each player's share should be proportional to their overall influence on various coalition structures.

The Shapley value's axiomatic properties make it a compelling and unique solution to fair allocation problems. It satisfies four main axioms: (1) efficiency, which ensures that the total value is evenly distributed among players; (2) symmetry, which indicates that players with identical contributions receive equal payoffs; (3) dummy player, which means that a player who contributes nothing to any coalition receives zero value; and (4) additivity, which allows the Shapley value to be applied to multiple games by summing individual contributions. These properties establish the Shapley value as the only allocation rule that accurately accounts for each player's contribution in a cooperative setting. However, its computational complexity poses a challenge, as calculating the exact Shapley value necessitates evaluating all possible coalitions, making it impractical for large-scale problems.

Despite its strong theoretical foundations, the practical application of the Shapley value is often hindered by its computational complexity. The exact computation necessitates evaluating all possible coalitions of players, resulting in an exponential increase in complexity with the number of participants [6]. Since the Shapley value entails high computational complexity, this may limit its applicability in real-time scenarios. To address this issue, researchers have created a variety of fast approximation methods, such as Monte Carlo sampling [7], linear regression approximations [3], and kernel-based approaches [8]. Furthermore, specialized algorithms for structured data have been proposed, such as characteristic functions with additivity constraints or situations in which contributions can be computed efficiently in closed form.

Aside from fast computation, there has been significant progress in extending the classical Shapley framework to new settings. These extensions include weighted and asymmetric Shapley values, interaction indices that capture player synergies, and adaptations for dynamic or time-evolving systems. Theoretical advancements have also investigated the relationship between Shapley values and other allocation principles, such as Aumann–Shapley pricing and the concept of cooperative fairness.

Shapley values have numerous applications in a variety of fields. They are a key component of explainable AI in machine learning [9], providing insights into model decisions via Shapley-based feature attributions [10]. In economics and social sciences, they aid in the measurement of power distributions in voting systems as well as economic contribution in collaborative networks. In data valuation, they are used to price data contributions in federated learning and data marketplaces.

This paper provides a structured survey of recent advances in Shapley value computation, extensions, and applications. The most recent methods for efficient computation, such as deterministic and stochastic approaches, are discussed. We then investigate key theoretical extensions that generalize the Shapley framework to more complex scenarios. Finally, we focus on practical applications in machine learning, economics, and data-driven decision-making. Through this survey, we aim to provide a comprehensive understanding of the current state of Shapley value research while also identifying open challenges for future research.

## 2. Fast Computation of Shapley Value

The computational complexity of Shapley values arises primarily from the exhaustive enumeration of all possible feature subsets. To compute Shapley values for a model with  $d$  features, the outputs must be evaluated across  $2^d$  distinct subsets. Direct computation is impractical for large-scale datasets and high-dimensional feature spaces due to its exponential complexity, particularly as the number of features increases. Such challenges are exacerbated in situations where efficiency and scalability are critical.

To address these issues, researchers have developed fast computation algorithms, which are broadly classified as model-agnostic approximation methods and model-specific approaches. This section provides an overview of these techniques, focusing on their fundamental principles, strengths, and limitations.

### 2.1. Model-Agnostic Approximation Algorithms

Model-agnostic approximation algorithms estimate Shapley values independent of the underlying model's structure, making them applicable across a wide range of machine learning frameworks. While these methods offer significant flexibility, their estimates often vary. This section discusses three prominent model-agnostic approaches, each of which uses a unique computational strategy. The detailed algorithms can be found in [11].

#### 2.1.1. Random Order Value

The Random Order Value algorithm (ROV) [11] estimates Shapley values by randomly permuting feature orders and calculating their marginal contributions. Based on the random order characterization of Shapley values, which defines them as the average marginal contribution across all feature permutations, the algorithm samples a fixed number of permutations  $m$  to estimate the Shapley value for a feature  $i$  as Equation (2):

$$\phi_i = \frac{1}{m} \sum_{k=1}^m \left[ v(S_{\pi_{k,i}} \cup \{i\}) - v(S_{\pi_{k,i}}) \right], \quad (2)$$

where  $\pi_k$  is a random permutation of features,  $S_{\pi_{k,i}}$  represents the set of features preceding feature  $i$  in the permutation  $\pi_k$ ,  $v(S)$  is the model output for subset  $S$ , and  $m$  represents the number of permutations sampled.

The Monte Carlo method (MC) [6] is a popular implementation of this approach, which generates random permutations and calculates marginal contributions. Adaptive sampling techniques are often employed to improve convergence by dynamically adjusting the sample size based on estimated variance. This algorithm is consistent with the probabilistic interpretation of Shapley values, as discussed in [11,12], which expresses the Shapley value as the expected marginal contribution across all feature permutations. By sampling a subset of permutations, the ROV algorithm provides a practical approximation of this theoretical concept.

The algorithm's key strengths are its model-agnostic nature and scalability, which reduce computational complexity from  $O(2^d)$  to  $O(d)$ . The Monte Carlo method, in particular, provides unbiased estimates while ensuring convergence to the true Shapley value as the number of samples increases. However, the approach has limitations, including high variance in estimates for large feature sets and significant computational costs from repeated model evaluations. Additionally, convergence can be slow, particularly when marginal contributions are highly variable. Future research could focus on more efficient sampling techniques, adaptive methods, or parallel computing frameworks to improve convergence, scalability, and cost-effectiveness.

### 2.1.2. Least Squares Value

The Least Squares Value algorithm (LSV) [13–15] calculates Shapley values by solving a weighted least squares problem. The Shapley value is defined as the coefficients of an additive model that minimizes the weighted squared error between model predictions and coalitional game values. The algorithm approximates the Shapley value by sampling a fixed number of subsets as Equation (3):

$$\phi_i = \operatorname{argmin}_{\beta} \sum_{S \subseteq D} W(S) \left[ v(S) - \left( \beta_0 + \sum_{j \in S} \beta_j \right) \right]^2, \quad (3)$$

where  $W(S) = (|D| - 1) / \left[ C_{|D|}^{|S|} |S| (|D| - |S|) \right]$  represents the weighting kernel. In this paper, we use  $C_b^a$  to denote  $b! / (a!(b - a)!)$  for short. In addition,  $v(S)$  is the coalitional game value for subset  $S$ ,  $\beta_0$  represents the intercept, and  $\beta_j$  is the coefficient corresponding to each feature.

A prominent implementation of this approach is the KernelSHAP method [2], which samples subsets of features and solves the weighted least squares problem. KernelSHAP is unbiased and asymptotically consistent, making it a popular technique for estimating Shapley values. Another variant, SGD-Shapley [16], uses stochastic gradient descent to iteratively solve the least squares problem, but it tends to be more biased than KernelSHAP.

The LSV algorithm is based on the least squares representation of the Shapley value, as discussed in [2] and further analyzed in [8]. This characterization interprets the Shapley value as the solution to a weighted least squares problem, with weights determined by subset size. The algorithm accurately approximates Shapley values by sampling subsets and solving the corresponding regression problem.

The algorithm's main advantages are its model-agnostic nature and theoretical guarantees of unbiasedness and consistency. KernelSHAP, in particular, provides a practical and efficient way to estimate Shapley values in complex models. However, the method has limitations, including the need for a large number of samples to obtain accurate estimates and the computational cost of solving the regression problem. Future research could concentrate on increasing convergence rates through more efficient sampling techniques or using parallel computing frameworks to improve scalability and lower computational costs.

### 2.1.3. Multilinear Extension Sampling

The multilinear extension sampling method (MES) [12] uses the cooperative game's multilinear extension to efficiently approximate Shapley values. This method extends the utility function to a continuous domain, allowing sampling techniques to estimate Shapley values without exhaustively evaluating all feature subsets [17]. The Shapley value for feature  $i$  is expressed using the multilinear extension as Equation (4):

$$\phi_i = \int_0^1 \mathbb{E}_{S_p \sim \text{Bernoulli}(p)} [v(S_p \cup \{i\}) - v(S_p)] dp, \quad (4)$$

where  $S_p$  is a random subset of features sampled according to a Bernoulli distribution with parameter  $p$ , and  $v(S_p)$  represents the utility function evaluated on subset  $S_p$ . The MES method approximates the integral by sampling values of  $p$  and subsets  $S_p$ . Specifically, for each sampled  $p$ , a subset  $S_p$  is generated by including each feature independently with a probability of  $p$ . The marginal contribution of feature  $i$  is calculated as  $v(S_p \cup \{i\}) - v(S_p)$ , and the Shapley value is estimated by averaging these contributions over multiple samples. This method is based on the multilinear extension of cooperative games, which yields a continuous representation of the utility function. As demonstrated in [17], this extension

allows for efficient sampling-based approximations of Shapley values, especially in high-dimensional settings where exact computation is computationally infeasible.

The MES method has significant advantages, including scalability and flexibility, because it can be applied to any utility function without requiring knowledge of the model's internal structure. However, its accuracy is dependent on the number of samples, and it may show high variance for complex utility functions. Future research could concentrate on improving sampling efficiency and lowering variance via adaptive sampling techniques or integration with parallel computing frameworks. Additionally, investigating its application in specific domains, such as health care or finance, could further demonstrate its practical utility.

## 2.2. Model-Specific Fast Algorithms

Model-specific fast algorithms are a specialized class of Shapley value estimation methods that use the inherent structure of specific models to significantly reduce computational complexity. These algorithms create efficient computation strategies for specific model types by leveraging the model's unique properties. This method not only provides high estimation accuracy but also significantly reduces computation time. Below, we outline several well-known model-specific fast algorithms.

### 2.2.1. Linear Models

Linear models are widely used in machine learning because they are simple and easy to interpret. The Shapley value, derived from cooperative game theory, has been applied to explain feature contributions in linear models. Refs. [18,19] provide fundamental insights into efficiently computing Shapley values for linear models, with an emphasis on their utility in model explanation and data valuation.

For a linear model  $f(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d + \epsilon$ , where  $\epsilon$  is the noise term,  $x_i$  is the feature value, and  $\beta_i$  is the coefficient of feature  $i$ , the Shapley value expression (LinearSHAP) [2,20] simplifies to Equation (5):

$$\phi_i(f, x^e) = \beta_i(x_i^e - \mu_i), \quad (5)$$

where  $x_i^e$  is the feature value of the explained sample, and  $\mu_i$  represents the mean of feature  $i$ . This formula shows that the Shapley value for linear models is determined solely by the feature coefficients and means, eliminating the need to enumerate all feature subsets, with a computational complexity of  $O(d)$ .

The LinearSHAP calculates Shapley values directly from the model's coefficients. This approach takes advantage of the model's linearity to avoid the computational complexity of permutation-based methods. The Shapley value for each feature is proportional to its weight in the model, which is scaled by the difference between the feature value and the expected value.

This method is based on the properties of linear models, in which feature contributions are additive and independent of other features. This enables a closed-form solution to the Shapley value, as demonstrated in [19]. The method is particularly efficient because it does not require sampling or approximation, making it appropriate for high-dimensional datasets.

The LinearSHAP has several advantages, including computational efficiency and exactness, because it computes feature contributions directly from the model's coefficients. However, it is only applicable to linear models and may not capture feature interactions in more complex models. Future research could focus on expanding this approach to nonlinear models or combining it with other Shapley value approximation techniques to

improve scalability and applicability. Furthermore, investigating its application in specific domains, such as finance or healthcare, could further demonstrate its practical utility.

### 2.2.2. Tree Models

Tree models, such as decision trees, random forests, and gradient-boosting trees, are popular in machine learning due to their interpretability and ability to handle nonlinear relationships. The Shapley value, derived from cooperative game theory, has been applied to explain feature contributions in tree models. Refs. [10,21] provide fundamental insights into efficiently computing Shapley values for tree models, with a focus on their utility in model explanation and data valuation. For tree-based models, the Shapley value for each feature is calculated using the TreeSHAP algorithm [10], which recursively assigns contributions based on the fraction of training samples that pass through each decision node, ensuring that the sum of contributions equals the difference between the model's output and the expected value.

The Interventional TreeSHAP method [10] computes Shapley values directly by leveraging the hierarchical structure of tree models. This method avoids the computational complexity of permutation-based methods by iteratively traversing the tree and calculating the contribution of each feature at each node. Specifically, the Shapley value for each feature is derived from the marginal contributions of the feature across all paths in the tree. This method is based on the properties of tree models, in which the contribution of each feature can be calculated recursively by traversing the trees. This enables an exact and efficient calculation of the Shapley value, as demonstrated in [10]. The method is especially efficient because it eliminates the need for sampling or approximation, making it suitable for high-dimensional datasets.

The Interventional TreeSHAP method has significant advantages, including computational efficiency and exactness, because it computes feature contributions directly from the tree's structure. However, it is only applicable to tree models and may fail to capture interactions between features in more complex models. Future research could look into extending this approach to hybrid models or integrating it with other Shapley value approximation techniques to improve scalability and applicability. Additionally, investigating its application in specific domains, such as finance or healthcare, could further demonstrate its practical utility.

### 2.2.3. Deep Models

Deep models, such as convolutional neural networks (CNNs) and multilayer perceptrons (MLPs), are widely used in machine learning because of their ability to capture complex patterns in high-dimensional data. However, explaining the contributions of features in deep models remains difficult due to their nonlinear and hierarchical architectures. DeepSHAP [2] and Gradient Shapley (G-Shapley) [22] are two well-known methods that address this problem by using Shapley values to provide interpretable feature attributions for deep models.

DeepSHAP and G-Shapley both use the concept of interventional Shapley values to quantify the contribution of each feature by comparing the model's output with and without the feature. These methods apply the Shapley value framework to deep models by using approximations that strike a balance between computational efficiency and accuracy. Deep models' Shapley values are typically approximated using methods such as DeepSHAP, which propagates contributions through network layers based on the DeepLIFT rescale rule, or G-Shapley, which integrates gradients along the path from a baseline input to the target input. There is no closed-form expression similar to tree models.

DeepSHAP extends the rescale rule from DeepLIFT to propagate Shapley values through each layer of a deep model [2]. For a deep model  $f(x) = (h_k \circ \dots \circ h_1)(x)$ , with  $h_i$  represents the  $i$ -th layer, and the Shapley value for layer  $i$  is calculated as:

$$\psi_i = \hat{\phi}(h_i, x^e, x^b) \cdot \frac{\psi_{i+1}}{f_i(x^e) - f_i(x^b)}, \quad (6)$$

where  $\psi_i$  represents the Shapley value at the  $i$ -th layer,  $h_i$  represents the  $i$ -th layer's operation,  $\hat{\phi}$  represents an approximation of the Shapley value for layer  $i$ ,  $f_i(x) = (h_i \circ h_{i-1} \circ \dots \circ h_1)(x)$  represents the compositional function of the first  $i$ -th layers,  $x^e$  is a explicanda sample, and  $x^b$  is a baseline sample.  $\hat{\phi}(h_i, x^e, x^b)$  represents an approximation of the Shapley values for the  $i$ -th layer operation  $h_i$ , computed via linearized local propagation rules to capture the contribution differences between the input  $x^e$  (explicand) and the baseline  $x^b$ , ensuring adherence to SHAP's efficiency and consistency axioms. This method ensures that attributions are efficient and equal to the difference between the model's output for the explicand and the baseline. DeepSHAP is especially useful for deep models because it eliminates the need for extensive sampling and permutation.

G-Shapley estimates Shapley values for deep models through gradient-based approximations [10]. The method calculates the gradient of the model's output in relation to each feature and integrates these gradients along a path from a baseline to the explicand. The Shapley value for feature  $i$  is approximated as:

$$\phi_i(f, x^e) = \int_0^1 \frac{\partial f(x^b + \alpha(x^e - x^b))}{\partial x_i} d\alpha, \quad (7)$$

where  $x^b$  represents the baseline sample, and  $\alpha$  parameterizes the path from the baseline to the explicand. G-Shapley is especially useful for models in which exact Shapley value computation is computationally infeasible, as it provides a scalable approximation.

DeepSHAP and G-Shapley provide efficient and interpretable methods for explaining deep models using Shapley values. DeepSHAP uses the rescale rule to propagate attributions through model layers, whereas G-Shapley uses gradient-based approximations to estimate feature contributions. Both methods are computationally efficient, but they may result in approximation bias, especially for highly nonlinear models. Future research could concentrate on reducing this bias and investigating their application in specific domains, such as health care or finance, to better demonstrate their practical utility. Additionally, developing more efficient sampling techniques or leveraging parallel computing frameworks could improve scalability while lowering computing costs.

To summarize the methods discussed above, we present a comprehensive comparison of model-agnostic approximation algorithms and model-specific fast algorithms for estimating Shapley value in Table 1. The table summarizes their key characteristics, such as strengths, limitations, and appropriate application scenarios. Model-agnostic algorithms are extremely versatile and can be applied to a wide range of machine learning models, making them ideal for situations requiring model flexibility. However, they often suffer from high computational complexity and variance, especially for large feature sets, which can cause slower convergence and higher computational costs. Future research could concentrate on developing more efficient sampling techniques, adaptive methods to reduce variance, and using parallel computing frameworks to improve scalability. Model-specific algorithms excel at computational efficiency and provide precise Shapley value estimates for specific model types, making them ideal for targeted applications. However, their applicability is limited to specific model structures, and they may struggle with complex or hybrid models involving significant feature interactions. Future directions include

extending these methods to hybrid or nonlinear models, improving bias reduction techniques, and investigating domain-specific applications (e.g., healthcare, finance) to increase their usefulness.

**Table 1.** Summary of model-agnostic approximation algorithms and model-specific fast algorithms.

Method	Applicable Models	Advantages	Disadvantages	Applicable Scenarios
ROV [11]	Model-agnostic	Unbiased estimation; adaptive sampling improves convergence	High variance; computationally expensive due to repeated model evaluations	Scenarios with few features and low model evaluation costs
KernelSHAP [2]	Model-agnostic	Low bias; asymptotically unbiased; antithetic sampling improves convergence	High computational complexity; requires many model evaluations	Moderate feature sets with manageable model evaluation costs
MES [12]	Model-agnostic	High computational efficiency; antithetic sampling improves convergence	High variance; computationally intensive for complex utility functions	High-dimensional feature spaces with low model evaluation costs
LinearSHAP [2,20]	Linear Models	Extremely efficient; provides exact Shapley values	Only applicable to linear models	Linear regression, logistic regression, and other linear models
Interventional TreeSHAP [10]	Tree Models	Highly efficient; exact computation; supports tree ensembles	Limited to tree-based models	Decision trees, random forests, gradient boosting trees
DeepSHAP [2]	Deep Models	Efficient; supports hybrid model interpretation	Potential bias; accuracy declines with highly complex model structures	Deep neural networks and hybrid models
G-Shapley [22]	Deep Models	Scalable; efficient for deep models; integrates gradients over paths	Potential bias; accuracy declines with highly nonlinear models	Deep neural networks, particularly in scenarios where exact computation is infeasible

Future advances in Shapley value computation can strike a balance between adaptability, scalability, and accuracy by combining the strengths of both approaches—flexibility from model-agnostic methods and efficiency from model-specific methods. This will make Shapley value estimation more practical in real-world applications, particularly with high-dimensional and large-scale datasets.

3. Extensions of the Shapley Value

The Shapley value is commonly used to quantify the contributions of individual data points when training machine learning models. Despite its foundational role, the traditional data Shapley framework has a significant limitation: it only provides values for points within a fixed dataset and does not take into account the data’s statistical properties. This limitation makes it challenging to reason about points outside of the dataset and ensure consistency when transferring data points between datasets. The distributional Shapley value proposed in this paper addresses these issues by defining a data point’s value in the

context of its underlying data distribution, resulting in a more stable and generalizable valuation method.

### 3.1. Distributional Shapley Value

To address the limitations of working with data from a fixed dataset, [23] extended the traditional Shapley value framework and introduced the distributional Shapley value. The proposed distributional Shapley value defines the value of a data point within the context of an underlying data distribution. This extension aims to provide a more stable and generalizable valuation method that can account for statistical aspects of the data while also reasoning about points outside the dataset. For a data point  $z$ , the distributional Shapley value is defined as the expected data Shapley value over datasets of size  $m$  drawn from the data distribution  $D$ , which is given as follows:

$$v_D(z; U, D, m) = E_{B \sim D^{m-1}}[\phi(z; U, B \cup \{z\})], \quad (8)$$

where  $\phi(z; U, B)$  denotes the traditional data Shapley value of point  $z$  in dataset  $B$  with a utility function  $U$ , and  $D^{m-1}$  indicates the distribution without  $z$ . In other words, the data Shapley value is treated as a random variable that depends on the specific draw of data from  $D$ . Taking the distributional Shapley value  $v_D(z; U, D, m)$  to be the expectation of this random variable eliminates the instability caused by the variance of  $\phi(z; U, B)$ . This approach ensures that the valuation remains stable in the face of perturbations in the data points and underlying data distribution.

To estimate the distributional Shapley values efficiently, [23] introduced the D-Shapley algorithm. This algorithm takes advantage of the distributive Shapley value's stability properties to provide precise estimates with formal guarantees. The algorithm involves sampling subsets of the data distribution and updating the estimates iteratively. Specifically, for each iteration  $t$ , a subset  $S_t$  is sampled from the distribution  $D$ , and the marginal contribution of each data point  $z$  is computed as  $\Delta_z U(S_t) = U(S_t \cup \{z\}) - U(S_t)$ . The estimates are then updated using the following equation:

$$v_{D,t+1}(z) = \frac{t-1}{t} \Delta_z U(S_t) + \frac{1}{t} v_{D,t}(z). \quad (9)$$

The algorithm generates unbiased estimates and provides formal guarantees of their quality.

### 3.2. Weighted Shapley Value

Beta Shapley [24] provides another extension to the traditional Shapley value framework. This extension seeks to address the limitations of the original Shapley approach, particularly in cases where uniform weighting of marginal contributions is not optimal. Ref. [24] argued that in machine learning settings, uniform averaging of marginal contributions, as done in data Shapley, can be suboptimal, particularly when the set's cardinality is high, resulting in negligible marginal contributions and making it difficult to distinguish the impact of individual data points. To address this, the paper relaxes the Shapley value's efficiency axiom, which states that the sum of individual data values must equal total utility. This relaxation enables a more flexible weighting of marginal contributions, resulting in the Beta Shapley formulation. For a data point  $z$ , the Beta Shapley value is expressed as a weighted sum of marginal contributions:

$$v_\beta(z; U, D) = \frac{1}{n} \sum_{j=1}^n \frac{1}{|D_j^{\setminus z}|} \sum_{S \in D_j^{\setminus z}} w_{\{\alpha, \beta\}}(j) [U(S \cup \{z\}) - U(S)], \quad (10)$$

where  $w_{\{\alpha, \beta\}}(j)$  are the normalized weights derived from the Beta distribution parameters  $\alpha$  and  $\beta$ ,  $D_j^{\setminus z} = \{S \subseteq D_j \setminus \{z\}, |S| = j - 1\}$ . These weights allow for emphasizing smaller or larger cardinalities depending on the choice of  $\alpha$  and  $\beta$ . For instance, when  $\alpha > \beta = 1$ , weights prioritize smaller cardinalities, resulting in improved performance in machine learning tasks, such as detecting mislabeled training data, learning with subsamples, and identifying influential data points.

The Beta Shapley framework brings together several popular data valuation methods under one umbrella and offers a more nuanced approach to quantifying the contribution of individual data points, adapting to various machine learning settings and data characteristics. To make the framework practical for large datasets, [24] proposed an efficient Monte Carlo algorithm for estimating Beta Shapley values. This algorithm selects subsets of varying sizes based on the Beta distribution weights and computes the marginal contributions, which are then averaged to estimate the Beta Shapley value. The empirical results show that Beta Shapley outperforms other methods, including data Shapley, on a variety of metrics and datasets, highlighting its effectiveness and utility in real-world scenarios.

### 3.3. Shapley Value for Reinforcement Learning

Based on reinforcement learning, [25] proposed a novel approach to explaining cooperative strategies and agent contributions in multiagent reinforcement learning (MARL) based on Shapley values, a game theory concept. Heuillet et al. [25] aimed to address the lack of explainability in deep reinforcement learning (DRL) models, particularly in multiagent systems where understanding each agent's contribution is critical for trust and transparency. The Shapley value was used to quantify each agent's contribution to the total reward, providing a fair and interpretable method for evaluating individual contributions. To manage computational complexity, [25] approximated the Shapley values using Monte Carlo sampling, as exact computation is infeasible due to the exponential growth of coalitions with agent number. The Shapley value for an agent  $i$  is calculated as follows:

$$v(i; U) = \frac{1}{|P|!} \sum_{S \subseteq P \setminus \{i\}} |S|!(|P| - |S| - 1)! [U(S \cup \{D_i\}) - U(S)], \quad (11)$$

where  $P$  represents the set of all agents,  $S$  is a subset of agents, and  $U(S)$  is the value function representing the expected global reward for the coalition  $S$ . In the reinforcement learning context, the Shapley value for agent  $i$  is calculated as the average marginal contribution of agent  $i$  across all agent permutations, ensuring that the global reward is distributed fairly to each agent. Ref. [25] demonstrated the effectiveness of this approach in two multiagent environments: Multiagent Particle and Sequential Social Dilemmas, demonstrating that Shapley values can accurately estimate each agent's contribution to the overall reward.

The paper further investigated the use of Shapley values to explain emergent behaviors in cooperative settings, emphasizing the importance of understanding agent contributions in improving the transparency and trustworthiness of MARL models. Ref. [25] also discussed Shapley values' limitations in providing detailed explanations for specific actions or episodes, emphasizing that they provide a high-level, average contribution metric.

### 3.4. Shapley Interaction Index

The Shapley Interaction Index has been a hot topic in the field of explainable artificial intelligence, with two notable papers influencing its development and application. Ref. [26] proposed the Shapley–Taylor index, which generalizes the Shapley value and assigns model predictions to interactions of subsets of features up to a certain size  $k$ . The method works similarly to the truncated Taylor Series, decomposing the function value at one point using its derivatives at another. The authors demonstrated that the Shapley–Taylor index

is equivalent to the Taylor Series of the multilinear extension of the model's set-theoretic behavior. For data point  $z$  in dataset  $B$ , the core formula for the Shapley–Taylor index is as follows:

$$I(z, U) = \frac{k}{n} \sum_{S \subseteq B \setminus \{z\}} \frac{U(S \cup \{z\}) - U(S)}{C_{n-1}^{|S|}}. \quad (12)$$

Ref. [26] axiomatized this method using the standard Shapley axioms—linearity, dummy, symmetry, and efficiency—and introduced an additional axiom known as the interaction distribution axiom. This new axiom describes how interactions are distributed across a class of functions that model pure interaction. The Shapley–Taylor index is demonstrated to meet these axioms and is unique in doing so. This work also provides a connection to the Taylor series of the multilinear extension, demonstrating that the Shapley–Taylor index can be derived from the Taylor series terms. Furthermore, [27] introduced a novel method for attributing feature interactions in machine learning models. The authors proposed the Faith–Shap index, which broadens the Shapley value to include interactions by taking the most faithful linear approximation to the pseudo-Boolean coalition game value function. The core idea is to define a general family of faithful interaction indices and then apply additional axioms to obtain a unique index. The interaction efficiency is the key axiom, which ensures that the sum of interaction weights equals the difference between the value function evaluated over the complete and empty sets. The Faith–Shap index is calculated by solving a weighted linear regression problem, allowing for efficient computation.

#### 4. Application

In cooperative game theory, various allocation methods are used to solve the problem of distributing benefits or costs among alliance members. For example, the Core method proposed by Gillies is defined as a set of allocations where no participant can obtain a higher benefit by leaving the alliance [28]. Its idea of alliance stability is that no sub-alliance can improve its own benefits through independent action, thereby ensuring global stability. The Gately point, proposed by Dermot Gately, determines the allocation scheme by minimizing the maximum dissatisfaction among alliance members (i.e., the motivation for members to leave the alliance) [29]. This method balances the interests of all parties and effectively reduces conflicts. Additionally, there is a Nash bargaining solution applicable to bilateral cooperation scenarios, which can maximize the product of participants' utilities [30]. However, in practical applications, compared to the Core and Gately point methods, the Shapley value has a strict unique solution and stronger interpretability. The Shapley value, with its fairness and scalability, is often used in cross-domain precise allocation problems and serves as an important standard for distributing total benefits among players.

The axiomatic definition of the Shapley value was first proposed by Lloyd Shapley in 1953 and is commonly used to determine fair and efficient resource allocation strategies within a group [1]. Subsequent research expanded into other fields, with Shapley and Shubik applying the Shapley value to voting games to propose the Shapley–Shubik index for quantifying the actual influence of voters [31]. As machine learning models become increasingly complex, the demand for model interpretability has grown, and Shapley values have emerged as an effective feature selection method. Strumbelj et al. first applied Shapley values to the field of machine learning, proposing the use of Shapley values to explain machine learning model predictions [32]. Ref. [2] provided a detailed introduction to the Shapley additive explanations framework for machine learning interpretability, and unified other explainability methods such as LIME and DeepLIFT under the same theoretical framework, thereby opening the door to the application of SHAP in the field of machine learning explainability.

The Shapley value's core idea is to allocate each feature's "contribution" to the overall prediction by taking into account all possible feature combinations. This ensures a fair distribution of the impact each feature has on a model's decision-making process, allowing the model's behavior to be understood even in complex scenarios. The Shapley value has been widely used in machine learning for model-agnostic interpretation, making it applicable to any type of model, including black-box models such as deep neural networks [2]. The method has also been used in feature selection, which helps identify which features are most influential in making predictions [32]. Additionally, Shapley values are increasingly being used in ensemble methods and algorithms to improve the fairness of machine learning models by providing a better understanding of how different input variables interact with one another when contributing to predictions. Despite their utility, the computational cost of calculating exact Shapley values can be prohibitively high for large datasets, leading to the development of approximate methods to make the use of Shapley values more practical [20]. Additionally, it has been used in data marketplaces to determine the value of data, allowing for fair pricing and trading of data resources [33]. Overall, the Shapley value remains a valuable tool for improving model interpretability, fairness, and transparency, and it has emerged as a key tool for interpreting and assigning values. In this section, we will go over its research and practical applications in various fields in greater detail, demonstrating its importance and the need for further research. It should also provide ideas for future Shapley development by researchers from various fields.

#### 4.1. Application in Health

Machine learning methods are currently being used extensively in health research to create data-driven, personalized, and resource-saving healthcare systems, as well as to improve diagnostic efficiency. For example, in the field of medical imaging, modern machine learning methods perform well in a variety of medical image analysis tasks and can accurately detect pneumonia in chest X-ray scans, significantly improving diagnostic results. Because training machine learning systems require large amounts of high-quality datasets, which necessitates accurate labeling of medical images, algorithms that automatically identify low-quality data are required to address this issue. Data valuation methods, an emerging field in AI research, can help to address the aforementioned challenges. Shapley values have emerged as an important indicator for determining data value in most recent studies because they uniquely satisfy the fundamental property of fair allocation. Therefore, in the health field, Shapley is often used to quantify the value of each training data pair for the performance of the predictor.

As previously stated, Shapley values are used to explain the predictions of deep learning models on medical images such as X-rays. For example, [34] used data Shapley values to assess the quality of training data for large chest X-ray datasets in the context of pneumonia detection. Beginning with the large public chest X-ray dataset ChestX-ray14, they used a pre-trained convolutional neural network (CNN) to extract features from the dataset, followed by the TMC-Shapley algorithm to calculate the value of each chest X-ray in pneumonia detection. Specifically, they recorded the training data as follows:

$$D = \{x_i, y_i\}_{i=1}^n, \quad (13)$$

where  $n$  represents the size of the training set,  $x_i$  is the feature vector for the  $i$ -th data point, and  $y_i \in \{0, 1\}$  represents the pneumonia label (0 and 1 indicate no and yes, respectively). Using a logistic regression algorithm, the prediction accuracy was used as a performance indicator to calculate the Shapley value of each training dataset and its relationship to the logistic regression algorithm's accuracy for pneumonia detection. This was used to assess the value of each chest X-ray. In general, low Shapley values indicated the presence of

mislabeled and low-quality images, whereas high Shapley values indicated the dataset's usefulness for pneumonia detection. Shapley's value was expressed as follows:

$$\phi_i = \sum_{S \subseteq D \setminus \{x_i\}} \frac{V(S \cup \{x_i\}) - V(S)}{C_{|D|-1}^{|S|}}, \quad (14)$$

where  $V(S)$  represents the accuracy of the pneumonia prediction on the validation set. The study discovered that removing training data with high Shapley values lowered pneumonia detection performance, while removing data with low Shapley values improved model performance. Therefore, the Shapley value can accurately quantify the importance of training data in the analysis of pneumonia cases.

Similarly, the extended data valuation method, which combines the Shapley value and the algorithm, will be more effective at valuing large amounts of medical image data. Ref. [35] investigated the feasibility of three different data valuation methods for medical image classification tasks and found that Shapley approximations based on k-nearest algorithms can handle large amounts of data examples in a reasonable time. The algorithm approximates complex deep neural network models using KNN classifiers, where a KNN classifier is trained for each deep feature in the test set. For a single labeled  $y_{val}$  query instance  $x_{val}$ , the KNN identifies the top K training data instances ( $x_{a_1}, \dots, x_{a_K}$ ) that are most similar to it, where similarity is defined by a specific distance metric. The corresponding labels of these instances ( $y_{a_1}, \dots, y_{a_K}$ ) then provide information about the label  $y_{val}$ . The confidence in correctly predicting the label is used as a performance score when calculating the KNN-Shapley value:

$$U(S) = \frac{1}{K} \sum_{k=1}^{\min(K, |S|)} 1[y_{a_k(s)} = y_{val}], \quad (15)$$

Shapley values are calculated recursively using:

$$\varphi(z_{a_N}) = \frac{1}{N} [y_{a_N} = y_{val}], \quad (16)$$

The Shapley value for the  $k$ th data example is calculated. Then, this formula is used:

$$\varphi(z_{a_i}) = \varphi(z_{a_{i+1}}) + \frac{1}{K} [y_{a_i} = y_{val}] - \frac{1}{K} [y_{a_{i+1}} = y_{val}] \frac{\min(K, i)}{i}, \quad (17)$$

Other Shapley values are calculated.

For 50,000 training data examples, the KNN-Shapley approximation takes approximately 24 h to calculate. It can effectively identify data examples that cause a drop in performance scores and prioritize the noise label verification process. It is a computationally efficient data valuation technique. The study demonstrated that Shapley, as a data evaluation method, can effectively assess the contribution of individual data examples when training ML models on real medical image datasets in a practical medical environment.

Since the introduction of the KernelSHAP method [2], Shapley values have become a standard metric for explaining machine learning model prediction results and improving model interpretability. Shapley values, particularly in clinical practice, can be used to assess the impact of each feature on model predictions, aid clinical decision-making, and thus broaden Shapley-based extensions such as median-SHAP to better explain black box models that predict human survival times [36]. The Shapley value can provide high predictability to the model, so its use in practical clinical applications warrants further investigation.

In addition, [37] demonstrated information-theoretic equivalence, indicating that Shapley values can be used for feature selection. Ref. [38] also proposed using Shapley

Additive Explanation to investigate the impact of diabetes indicator results. Shapley Additive Explanation is a method for screening redundant attributes in machine learning models that can accurately predict diabetes characteristics [39].

#### 4.2. Application in Finance

Shapley values are becoming more widely used in the financial industry, providing new perspectives and methods for risk management, revenue distribution, and market analysis. Shapley values can be used not only for cost allocation in companies [40], the valuation of corporate voting rights [41], and so on, but also to provide financial institutions with a fair and reasonable mechanism for evaluating and allocating contributions from all parties. In particular, for the cooperation of financial institutions, if financial institutions cooperate to some extent in non-core cooperation areas and generate synergies in regulatory reporting, the Shapley value can be used as a fair interpretation scheme to allocate the marginal contribution to such synergies. It considers the impact of individual institutional agents on the joint results of each institution, ensuring that financial institutions achieve cooperative transformation [42]. The application of the Shapley value provides strong support for the financial industry's continued development.

Shapley values can be used in equity risk management to quantify a security asset's relative risk within an optimal portfolio. Investors can determine the exact contribution of each risky asset to joint returns. Refs. [43,44] applied Shapley's value theory to price the market risk of individual assets. On this basis, [45] extended to optimal portfolios by calculating the Shapley value of the securities in the optimal portfolios and applying it to estimate the systematic risk of individual stocks. The study took the minimization of risk in portfolio selection as an objective, and the selection of appropriate portfolio weights minimized the portfolio variance  $\sigma_p^2$ , which was expressed as follows:

$$\sigma_p^2 = w' \Sigma w, \quad (18)$$

subject to  $w'l = 1$ , where  $w$  represents the portfolio weight array,  $\Sigma$  is the covariance matrix of each stock, and  $l$  represents the array of elements equal to 1.

The reduction in risk and increase in return was determined by the order in which the stocks were added to the portfolio, so the Sharpe ratio was calculated by averaging the marginal contributions of the stocks added to a particular portfolio:

$$SV_k(N, v) = \sum_{s=1}^{N-1} \sum_S \frac{s!(N-s-1)!}{N!} [v(S+k) - v(s)], \quad (19)$$

where  $s = 1, \dots, N$  is the number of stocks in the portfolio, and  $v(s)$  represents the risk associated with the optimal portfolio consisting of  $s$  stocks. In the study, the Shapley values of stocks and indices were calculated for the optimal mean-variance portfolio using a portfolio consisting of 13 stock and industry indices from 2016 to 2019 and daily adjusted returns. The traditional beta coefficient was calculated as follows:

$$\beta_i = \text{cov}(r_i, r_m) / \sigma_m^2, \quad (20)$$

When compared with the Shapley value, it was found that the Shapley value assigned the relative risk and return of other assets in the portfolio, which can more accurately predict the actual impact of the transaction, demonstrating the contribution of Shapley value theory in financial optimization.

In addition to stock risk allocation, Shapley values can be used to attribute a portfolio's actual performance to individual characteristics, providing investors with a reference point

for developing portfolio strategies. Ref. [46] proposed a Shapley values-based method for attributing portfolio performance to individual characteristics and discussed how to approximate this attribution. The Shapley attribution was the average of the lifted order-of-arrival attributions for all  $n$  permutations minus the baseline, which was repeated for each feature.

Vector  $x = (x_1, \dots, x_n)$  was used to represent the investment characteristic configuration value,  $f(x_1, \dots, x_n)$  evaluated the performance under the assumed configuration, the baseline value  $b = f(0)$  was set, and the Shapley attribution of the characteristics was as follows:

$$a_i = \left\{ \sum_{x \in X_i} \frac{(1^T x)!(n - 1^T x - 1)!}{n!} [f(x + e_i) - f(x)] \right\} - b, \quad (21)$$

where  $X_i$  is the set of configurations when the characteristic  $i$  is off. Using this Shapley attribution formula, the value of all investment characteristic configurations can be calculated directly. Monte Carlo sampling is an approximation method that can quickly approximate Shapley attribution.

$$a_i = E[f(x + e_i) - f(x)] - b, \quad (22)$$

where

$$\text{Prob}(x = x') = \frac{(1^T x)!(n - 1^T x - 1)!}{n!}. \quad (23)$$

Using the S&P 500 index as a benchmark portfolio with data spanning from 2002 to 2019, a simulation experiment was conducted to compare the attribution results of the Sharpley ratio method, the one-step method, and the leave-out method. It was discovered that the Sharpley ratio method had a better estimation effect and provided a new reference method for investment selection in the financial industry.

Its applications include analyzing fairness issues. The cohort Shapley value proposed by [47] as an improvement has the advantage of avoiding the extrapolation problem in Shapley value calculation, as well as model evaluation of infeasible combinations. The relative fairness score based on the cohort Shapley value can be used to calculate the degree of privilege or disadvantage of different groups in terms of factors like mobility, employment size, and company age. The calculation of this score can further assess the fairness of SMEs' access to external financing [48]. It is feasible for financial institutions to develop a relative fairness value method based on the Shapley value to assess the fairness of their credit decisions, which could become a new research direction.

#### 4.3. Application in Industry

The core strengths of Shapley values are their fairness and interpretability, which allow them to balance the demands of multiple stakeholders. Shapley values, as a powerful measurement model method, can quantify each participant's contributions to multi-party collaboration to solve complex problems such as efficiency optimization and responsibility allocation. In the field of personal production, Shapley values can quantify individual productivity in the service industry by treating hourly income as each employee's expected marginal contribution [49]. Shapley value regression methods are used in industrial machine production to evaluate the impact of various predictors on process performance, such as carbide furnace output [50]. In particular, for industrial environmental issues, the DEA–Game model calculates the Shapley values for each decision-making unit alliance, yielding the final ranking of industrial producers in terms of environmental efficiency [51]. It can also be used to allocate emission responsibilities using the Shapley values of cooperative games, allowing total carbon emissions to be redistributed among supply chain

companies [52]. In general, Shapley values not only increase industrial efficiency and quality but also promote the achievement of sustainable development objectives.

More importantly, Shapley values can be used to quantify profit distribution issues, such as the benefits to each party in a procurement network [53] or the profit distribution game in the supply chain. Ref. [54] started with the supply chain and investigated how to use Shapley values to allocate the expected excess profits generated by the inventory pooling effect among retailers and suppliers. Assuming  $\pi_{R_i}$  represents the expected profit of the retailer before centralized management and  $\pi_{S_i}$  represents the expected profit of the supplier retaining independent inventory for the supplying retailer  $i$ , in the inventory pooling game between  $N$  retailers and suppliers, the Shapley values of the retailer  $i$  and supplier  $s$  were allocated as follows:

$$\phi_{R_i} = \sum_{\substack{J \subseteq N \cup \{S\} \setminus \{i\} : \\ \{S\} \in J, |J| \geq 1}} \frac{(|J|)!(N - |J|)!}{(N + 1)!} \pi(J \cup \{i\}) - \sum_{\substack{J \subseteq N \cup \{S\} \setminus \{i\} : \\ \{S\} \in J, |J| \geq 2}} \frac{(|J|)!(N - |J|)!}{(N + 1)!} \pi(J) - \frac{1}{2}(\pi_{S_i} + \pi_{R_i}), \quad (24)$$

$$\phi_S = \sum_{\substack{J \subseteq N : \\ |J| \geq 1}} \frac{(|J|)!(N - |J|)!}{(N + 1)!} \pi(J \cup \{S\}) - \frac{1}{2} \sum_{i \in N} (\pi_{S_i} + \pi_{R_i}), \quad (25)$$

Allocation based on Shapley values can encourage suppliers to make the best inventory decisions for collective alliances. If all parties agree on the expected benefits of inventory centralization, resulting in a Shapley value for each participant, retailers will agree to adopt a shared inventory policy, and suppliers will carry the appropriate amount of inventory for supply chain benefits. Overall, the Shapley value is stable and reasonable in industrial profit distribution applications.

In today's rapidly evolving network technology, the Industrial Internet of Things is gradually merging with artificial intelligence, improving the operating model of manufacturing enterprises while raising concerns about communication costs. The integration of the Industrial Internet of Things and federated learning can effectively solve these problems, while the integration of the model based on Shapley value means it can identify industries with high-quality data partitions to add to the model and improve model performance even further. Ref. [55] combined Shapley values with a highly effective global training method for the Industrial Internet of Things. The contribution was quantified and aggregated using the Shapley values of the local industries that participate in the federated learning system, and aggregation weights were assigned based on dataset size and data heterogeneity to efficiently train the model. The objective of the study was to achieve high accuracy by minimizing prediction loss while reducing the computational cost of calculating Shapley values by quantifying collaborative contribution in each training round. The utility function  $v(M)$  used in the Shapley value calculation represented the accuracy of the central model trained using sub-clusters  $M$ .

In summary, Shapley values have enormous potential for use in industry. They promote collaborative industry development by providing effective allocation values, as well as theoretical support and practical guidelines for academia and industry.

#### 4.4. Application in Digital Economy

The digital economy is an economic activity based on digital technology, with data serving as the primary production factor, resulting in the rapid growth of the digital economy industry. Recent research has investigated the commercialization of data, resulting in the concepts of data pricing and data markets. Shapley values can play a significant role in the digital economy, addressing issues, such as data value distribution by quantifying

the contributions of data, algorithms, and specific behaviors on online platforms. At the same time, in the context of the rapid development of the digital economy, the efficiency improvement of various industries depends on the joint action of multiple factors. Shapley values can reasonably calculate the impact of each factor. In practical application scenarios, Shapley values can attribute the contribution of online advertising in the digital field [56], enabling advertisers to better understand their customers. Combining it with machine learning algorithms can also improve the effect. Using artificial neural networks and Shapley additive explanations, researchers can mine data on the nonlinear correlations and characteristics of the digital economy and energy productivity [57], allowing them to gain a better understanding of the complex relationship between the two.

Data pricing is the pricing of data as an asset, allowing it to be sold or purchased as a commodity in commercial and economic activities. It is typically based on the evaluation of data based on its characteristics and is one of the most important fields in the digital economy. In recent research, Shapley values have been used in this field to evaluate the contribution of data records to a model and to complete data pricing. Most of these studies have focused on using Shapley values as an indicator to quantify the contribution of individual data sources [58]. Kleinberg et al. [59] were the first to use Shapley values for private data pricing and investigated their use in marketing surveys, collaborative filtering, and recommendation systems. To ensure that buyers paid exactly the same amount as the data has been valued, as well as to meet the growing demand for secure data transactions in the data market, [60] developed the algorithm and proposed a Shapley value algorithm implemented through multi-party computation (MPC) that could effectively achieve fair payment in the data market.

In addition, [61] introduced Shapley values to measure the value of data in terms of fairness when considering the three-agent data market problem of data owners, model buyers, and brokers common in the digital economy, and constructed a revenue optimization problem based on the sum of Shapley values of data boundaries to obtain an optimal solution:

$$SV_i = \sum_{s \subseteq \{z_1, \dots, z_n\} \setminus z_i} \frac{V(s \cup \{z_i\}) - V(s)}{C_{|N|-1}^{|s|}}. \quad (26)$$

Then, the  $SV_i$  values are sorted, and  $S_{sv_n}$  denotes the sum of Shapley values given by the top  $n$  data, which is  $S_{sv_n} = \sum_{i=1}^n SV_i = \beta_1 - \beta_2 \exp(-\beta_3 n)$ . This indicates the accuracy level that the model can reach after adding  $n_{th}$  data record. Substituting it into the profit function and taking the partial derivative yielded the optimal data boundaries and data subscription fees, demonstrating that Shapley values can be used to design an efficient data transaction process.

Shapley values provide a fair and interpretable framework for data pricing. To further illustrate and value the utility generated by individuals in the digital economy, [62] considered users as fundamental components in the data value chain and proposed a method based on Shapley value approximation to estimate the fair compensation that each user should receive when providing ratings to the service's recommendation system, ensuring that users received a reward proportional to their contribution. The approximate Shapley value was calculated by clustering each user. After determining the Shapley value of each cluster, the cluster's centroid was marked with the corresponding Shapley value. The Shapley value of each user was the sum of all independent cluster values divided by the Euclidean distance between the user's point and the centroid of the corresponding group, plus a stable value. Finally, the assigned value was scaled based on the utility generated by the user across the entire dataset to ensure that the efficiency condition is satisfied. Assigning approximate Shapley values in this manner significantly simplified the calculation process.

In conclusion, in the digital economy, Shapley values not only ensure fairness in the benefits that each party derives from the data, but they can also be used to calculate the value that different users can provide for the data. This promotes efficient data resource allocation and has significantly enriched digital economy theories.

## 5. Conclusions

In this paper, we provide a comprehensive review of the most recent advances in the computation and extension of the Shapley value. We begin by explaining its fundamental definition and the four core axioms that underpin fair allocation. To address the inherent computational challenges posed by the need to evaluate all possible coalitions, we discuss both model-agnostic approximation techniques (such as ROV, LSV, and MES) and model-specific fast algorithms designed for linear models, tree-based models, and deep neural networks. Furthermore, we investigate several extensions to the classical Shapley framework, including Distributional Shapley, Weighted Shapley, and Shapley Interaction Indices, which broaden its application to data valuation, reinforcement learning, and feature interaction analysis. The Shapley value finds appropriate application in domains such as game theory and cooperative games, where it provides a robust framework for fair allocation among players based on their marginal contributions. It is particularly valuable in scenarios requiring precise attribution of contributions, such as machine learning feature importance analysis, economic cost-sharing problems, and situations demanding equitable distribution mechanisms. However, due to its computational complexity, which grows exponentially with the number of players, the Shapley value may not be suitable for real-time applications or large-scale systems where rapid decision-making is essential. Similarly, in contexts with an extremely high number of participants or where approximate solutions are acceptable, alternative methods with lower computational demands might be more appropriate despite potentially sacrificing some of the Shapley value's desirable theoretical properties.

Looking ahead, as the era of big data and complex modeling evolves, improving the computational efficiency of Shapley value estimation while maintaining accuracy remains a pressing research challenge. In particular, emerging areas such as data asset pricing and data trading present promising opportunities for interdisciplinary research, with the integration of Shapley value methodologies with financial econometrics and economic theory potentially leading to more precise data valuation and resource allocation. Further research into interpreting deep and hybrid models, as well as ensuring fairness in multi-agent decision-making, is expected to increase the practical utility of Shapley value approaches.

**Author Contributions:** Conceptualization, L.Q.; methodology, L.Q. and Y.Z. (Yingqiu Zhu); validation, L.Q., Y.Z. (Yingqiu Zhu) and Y.Z. (Yining Zhao); formal analysis, L.Q., Y.Z. (Yingqiu Zhu) and Y.Z. (Yining Zhao); investigation, Y.Z. (Yingqiu Zhu), S.L. and X.Z.; writing—original draft preparation, L.Q., Y.Z. (Yingqiu Zhu), S.L. and X.Z.; writing—review and editing, L.Q., Y.Z. (Yingqiu Zhu) and Y.Z. (Yining Zhao); supervision, L.Q. and Y.Z. (Yining Zhao) All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Major Project of National Social Science Foundation of China (No. 22&ZD072), and the National Natural Science Foundation of China (No. 72301070).

**Data Availability Statement:** No new data were created.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Shapley, L.S. *A Value for N-Person Games. Contributions to the Theory of Games*; Princeton University Press: Princeton, NJ, USA, 1953.
- Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
- Roth, A.E. The Shapley value as a von Neumann-Morgenstern utility. *Econometrica* **1977**, *45*, 657–664. [\[CrossRef\]](#)
- Jackson, M.O. Allocation rules for network games. *Games Econ. Behav.* **2005**, *51*, 128–154. [\[CrossRef\]](#)
- Young, H.P. Monotonic solutions of cooperative games. *Int. J. Game Theory* **1985**, *14*, 65–72. [\[CrossRef\]](#)
- Castro, J.; Gómez, D.; Tejada, J. Polynomial calculation of the Shapley value based on sampling. *Comput. Oper. Res.* **2009**, *36*, 1726–1730. [\[CrossRef\]](#)
- Mitchell, R.; Cooper, J.; Frank, E.; Holmes, G. Sampling permutations for Shapley value estimation. *J. Mach. Learn. Res.* **2022**, *23*, 1–46.
- Covert, I.; Lee, S.I. Improving KernelSHAP: Practical Shapley value estimation using linear regression. In Proceedings of the International Conference on Artificial Intelligence and Statistics. PMLR, Virtual, 13–15 April 2021.
- Ning, Y.; Ong, M.E.H.; Chakraborty, B.; Goldstein, B.A.; Ting, D.S.W.; Vaughan, R.; Liu, N. Shapley variable importance cloud for interpretable machine learning. *Patterns* **2022**, *3*, 100452. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [\[CrossRef\]](#)
- Monderer, D.; Samet, D. Variations on the Shapley value. *Handb. Game Theory Econ. Appl.* **2002**, *3*, 2055–2076.
- Winter, E. The Shapley value. *Handb. Game Theory Econ. Appl.* **2002**, *3*, 2025–2054.
- Charnes, A.; Golany, B.; Keane, M.; Rousseau, J. Extremal principle solutions of games in characteristic function form: Core, Chebychev and Shapley value generalizations. *Econom. Plan. Effic.* **1988**, 123–133.
- Ruiz, L.M.; Valenciano, F.; Zarzuelo, J.M. The family of least square values for transferable utility games. *Games Econ. Behav.* **1998**, *24*, 109–130. [\[CrossRef\]](#)
- Chen, H.; Covert, I.C.; Lundberg, S.M.; Lee, S.-I. Algorithms to estimate Shapley value feature attributions. *Nat. Mach. Intell.* **2023**, *5*, 590–601. [\[CrossRef\]](#)
- Simon, G.; Vincent, T. A projected stochastic gradient algorithm for estimating Shapley value applied in attribute importance. In *Machine Learning and Knowledge Extraction: International Cross-Domain Conference, Dublin, Ireland, 25–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020.
- Owen, G. Multilinear extensions of games. *Manag. Sci.* **1972**, *18*, 64–79. [\[CrossRef\]](#)
- Jia, R.; Dao, D.; Wang, B.; Hubis, F.A.; Hynes, N.; Gürel, N.M.; Li, B.; Zhang, C.; Song, D.; Spanos, C.J. Towards efficient data valuation based on the Shapley value. In Proceedings of the International Conference on Artificial Intelligence and Statistics. PMLR, Naha, Japan, 16–18 April 2019.
- Sundararajan, M.; Najmi, A. The many Shapley values for model explanation. In Proceedings of the International Conference on Machine Learning. PMLR, Virtual, 13–18 July 2020.
- Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [\[CrossRef\]](#)
- Amoukou, S.I.; Salaün, T.; Brunel, N. Accurate Shapley values for explaining tree-based models. In Proceedings of the International Conference on Artificial Intelligence and Statistics. PMLR, Virtual, 28–30 March 2022.
- Ancona, M.; Oztireli, C.; Gross, M. Explaining deep neural networks with a polynomial time algorithm for Shapley value approximation. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 9–15 June 2019.
- Ghorbani, A.; Kim, M.; Zou, J. A distributional framework for data valuation. In Proceedings of the International Conference on Machine Learning. PMLR, Virtual, 13–18 July 2020.
- Kwon, Y.; Zou, J. Beta Shapley: A unified and noise-reduced data valuation framework for machine learning. In Proceedings of the International Conference on Artificial Intelligence and Statistics. PMLR, Virtual, 28–30 March 2022.
- Heuillet, A.; Couthouis, F.; Díaz-Rodríguez, N. Collective explainable AI: Explaining cooperative strategies and agent contribution in multiagent reinforcement learning with Shapley values. *IEEE Comput. Intell. Mag.* **2022**, *17*, 59–71. [\[CrossRef\]](#)
- Sundararajan, M.; Dhamdhere, K.; Agarwal, A. The Shapley Taylor interaction index. In Proceedings of the International Conference on Machine Learning. PMLR, Virtual, 13–18 July 2020.
- Tsai, C.P.; Yeh, C.K.; Ravikumar, P. Faith-shap: The faithful Shapley interaction index. *J. Mach. Learn. Res.* **2023**, *24*, 1–42.
- Gillies, D.B. Solutions to general non-zero-sum games. *Contrib. Theory Games* **1959**, *4*, 47–85.
- Gately, D. Sharing the gains from regional cooperation: A game theoretic application to planning investment in electric power. *Int. Econ. Rev.* **1974**, *15*, 195–208. [\[CrossRef\]](#)
- Nash, J.F. The bargaining problem. *Econometrica* **1950**, *18*, 155–162. [\[CrossRef\]](#)
- Shapley, L.S.; Shubik, M. A method for evaluating the distribution of power in a committee system. *Am. Political Sci. Rev.* **1954**, *48*, 787–792. [\[CrossRef\]](#)

32. Strumbelj, E.; Kononenko, I. An efficient explanation of individual classifications using game theory. *J. Mach. Learn. Res.* **2010**, *11*, 1–18.
33. Chen, L.; Wang, H.; Chen, L.; Koutris, P.; Kumar, A. Demonstration of nimbus: Model-based pricing for machine learning in a data marketplace. In Proceedings of the 2019 International Conference on Management of Data, Amsterdam, The Netherlands, 30 June–5 July 2019.
34. Tang, S.; Ghorbani, A.; Yamashita, R.; Rehman, S.; Dunnmon, J.A.; Zou, J.; Rubin, D.L. Data valuation for medical imaging using Shapley value and application to a large-scale chest X-ray dataset. *Sci. Rep.* **2021**, *11*, 8366. [[CrossRef](#)] [[PubMed](#)]
35. Pandl, K.D.; Feiland, F.; Thiebes, S.; Sunyaev, A. Trustworthy machine learning for health care: Scalable data valuation with the Shapley value. In Proceedings of the Conference on Health, Inference, and Learning, Virtual, 8–10 April 2021.
36. Ter-Minassian, L.; Ghalebikesabi, S.; Diaz-Ordaz, K.; Holmes, C. Explainable AI for survival analysis: A median-SHAP approach. *arXiv* **2024**, arXiv:2402.00072.
37. Covert, I.; Lundberg, S.; Lee, S.I. Explaining by removing: A unified framework for model explanation. *J. Mach. Learn. Res.* **2021**, *22*, 1–90.
38. Chang, V.; Ganatra, M.A.; Hall, K.; Golightly, L.; Xu, Q.A. An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. *Healthc. Anal.* **2022**, *2*, 100118. [[CrossRef](#)]
39. Ejayi, C.J.; Qin, Z.; Amos, J.; Ejayi, M.B.; Nnani, A.; Ejayi, T.U.; Agbesi, V.K.; Diokpo, C.; Okpara, C. A robust predictive diagnosis model for diabetes mellitus using Shapley-incorporated machine learning algorithms. *Healthc. Anal.* **2023**, *3*, 100166. [[CrossRef](#)]
40. Lemaire, J. An application of game theory: Cost allocation. *ASTIN Bull. J. IAA* **1984**, *14*, 61–81. [[CrossRef](#)]
41. Zingales, L. What determines the value of corporate votes? *Q. J. Econ.* **1995**, *110*, 1047–1073. [[CrossRef](#)]
42. Janowski, A.; Plenk, M.; Haselwander, M. Sharing Economy in the Financial Industry: A Platform Approach towards Sharing in Regulatory Reporting using the Shapley Value. *SUERF Policy Note* **2021**, *225*, 1–14.
43. Ortmann, K.M. The link between the Shapley value and the beta factor. *Decis. Econ. Financ.* **2016**, *39*, 311–325. [[CrossRef](#)]
44. Colini-Baldeschi, R.; Scarsini, M.; Vaccari, S. Variance allocation and Shapley value. *Methodol. Comput. Appl. Probab.* **2018**, *20*, 919–933. [[CrossRef](#)]
45. Shalit, H. Using the Shapley value of stocks as systematic risk. *J. Risk Financ.* **2020**, *21*, 459–468. [[CrossRef](#)]
46. Moehle, N.; Boydb, S.; Angc, A. Portfolio performance attribution via Shapley value. *J. Invest. Manag.* **2022**, *20*, 33–52.
47. Mase, M.; Owen, A.B.; Seiler, B. Explaining black box decisions by Shapley cohort refinement. *arXiv* **2019**, arXiv:1911.00467.
48. Lu, X.; Calabrese, R. The Cohort Shapley value to measure fairness in financing small and medium enterprises in the UK. *Financ. Res. Lett.* **2023**, *58*, 104542. [[CrossRef](#)]
49. Rico Lugo, S.D.; Wu, D.; Nishino, N. Utilizing the Shapley Value to Measure Individual Productivity in the Service Industry. In Proceedings of the IFIP International Conference on Advances in Production Management Systems, Chemnitz, Germany, 8–11 September 2024; Springer Nature Switzerland: Cham, Switzerland, 2024.
50. Liu, X.; Aldrich, C. Assessing the influence of operational variables on process performance in metallurgical plants by use of Shapley value regression. *Metals* **2022**, *12*, 1777. [[CrossRef](#)]
51. Omrani, H.; Amini, M.; Babaei, M.; Shafaat, K. Use Shapley value for increasing power distinguish of data envelopment analysis model: An application for estimating environmental efficiency of industrial producers in Iran. *Energy Environ.* **2020**, *31*, 656–675. [[CrossRef](#)]
52. Gopalakrishnan, S.; Granot, D.; Granot, F.; Sošić, G.; Cui, H. Incentives and emission responsibility allocation in supply chains. *Manag. Sci.* **2021**, *67*, 4172–4190. [[CrossRef](#)]
53. Ryan, J.K.; Shao, L.; Sun, D. Contracting mechanisms for stable sourcing networks. *Manuf. Serv. Oper. Manag.* **2022**, *24*, 2558–2576. [[CrossRef](#)]
54. Kemahlioglu-Ziya, E.; Bartholdi, J.J., III. Centralizing inventory in supply chains by using Shapley value to allocate the profits. *Manuf. Serv. Oper. Manag.* **2011**, *13*, 146–162. [[CrossRef](#)]
55. Bhatti, D.M.S.; Ali, M.; Yoon, J.; Choi, B.J. Efficient Collaborative Learning in the Industrial IoT Using Federated Learning and Adaptive Weighting Based on Shapley Values. *Sensors* **2025**, *25*, 969. [[CrossRef](#)] [[PubMed](#)]
56. Singal, R.; Besbes, O.; Desir, A.; Goyal, V.; Iyengar, G. Shapley meets uniform: An axiomatic framework for attribution in online advertising. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019.
57. Sun, C.; Xu, M.; Wang, B. Deep learning: Spatiotemporal impact of digital economy on energy productivity. *Renew. Sustain. Energy Rev.* **2024**, *199*, 114501. [[CrossRef](#)]
58. Jia, R.; Dao, D.; Wang, B.; Hubis, F.A.; Gurel, N.M.; Li, B.; Zhang, C.; Spanos, C.J.; Song, D. Efficient task-specific data valuation for nearest neighbor algorithms. *Proc. VLDB Endow.* **2019**, *12*, 1610–1623. [[CrossRef](#)]
59. Kleinberg, J.; Papadimitriou, C.H.; Raghavan, P. On the value of private information. In Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge, Certosa di Pontignano, Italy, 8–10 July 2001; pp. 249–257.
60. Tian, Z.; Liu, J.; Li, J.; Cao, X.; Jia, R.; Kong, J.; Liu, M.; Ren, K. Private data valuation and fair payment in data marketplaces. *arXiv* **2022**, arXiv:2210.08723.

61. Tian, Y.; Ding, Y.; Fu, S.; Liu, D. Data boundary and data pricing based on the shapley value. *IEEE Access* **2022**, *10*, 14288–14300. [\[CrossRef\]](#)
62. Paraschiv, M.; Laoutaris, N. Valuating user data in a human-centric data economy. *arXiv* **2019**, arXiv:1909.01137.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.