Data Valuation and Pricing in Internet of Things: Survey and Vision

Xinyi Shi[123†], Haihan Duan[12*]
[1]*Artificial Intelligence Research Institute, Shenzhen MSU-BIT University*, Guangdong, China
[2]*Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Shenzhen MSU-BIT University*, Guangdong, China
[3]*School of Science and Engineering*, *The Chinese University of Hong Kong, Shenzhen*, Guangdong, China
xinyishi11@link.cuhk.edu.cn, duanhaihan@smbu.edu.cn

***Abstract*—In the digital age, data has become an invaluable asset for decision-making across various industries. Accurate data valuation is essential for businesses to effectively leverage their data assets, optimize strategies, and enhance operational efficiencies. This paper discusses the complex challenges inherent in data valuation methods, focusing on issues of data provenance and the lack of standardized valuation metrics. Moreover, the Internet of Things (IoT) further complicates this landscape by generating vast volumes of real-time data, which requires robust evaluation frameworks. Blockchain technology, with its decentralized and tamper-resistant characteristics, offers promising solutions by ensuring data integrity and traceability. Additionally, smart contracts enable automated and reliable execution of data transactions, reinforcing trust in the data exchange process. Technically, the integration of wireless sensing technology and edge computing facilitates real-time data collection and processing, improving the accuracy and timeliness of data valuation, and machine learning (ML) techniques further enhance these efforts by uncovering patterns and relationships within large datasets. This study explores how these advanced technologies can address the limitations of existing data valuation methods, paving the way for a more transparent, secure, and efficient data marketplace.**

***Index Terms*—Data Valuation, Pricing, Internet of Things, Blockchain**

I. Introduction

In the digital age, data has become a pivotal asset driving decision-making processes across various industries. The ability to accurately assess data is crucial for businesses to effectively utilize their data assets and cultivate a dynamic data marketplace. In the realm of data-driven innovation, data valuation—the process of assessing and quantifying the value of data—has emerged as a critical area of focus, which is especially important in the Internet of Things (IoT), since vast volumes of data generated by IoT devices can significantly enhance system intelligence and automation. Moreover, accurate data valuation can help organizations optimize resource allocation, improve data utilization efficiency, and drive innovative applications. However, the effective valuation and management of the data remain challenging in the current stage [1], [2].

Existing research in data valuation is varied and fragmented, encompassing economic models for market valuation and algorithms for quality and contribution assessment. Therefore, there is a notable gap in comprehensive analyses specific to IoT, despite the existing literature reviews, particularly in how data valuation integrates with technologies like blockchain, wireless sensing, machine learning (ML), and edge computing.

Beyond IoT, these insights into data valuation are relevant across healthcare, finance, and smart cities, where data-driven decision-making is paramount. Healthcare organizations can use robust data valuation frameworks to optimize patient outcomes, while financial institutions refine risk management by leveraging more precise data assessments. Similarly, smart city infrastructure can benefit from real-time data valuation to enhance public services and resource management, while sectors like retail and manufacturing can leverage data to improve supply chains and customer experiences.

Existing surveys, such as those conducted by Sen *et al.* [1] and Liang *et al.* [2], have summarized various aspects of data valuation but exhibit significant and notable limitations. They fail to discuss the unique characteristics and challenges of IoT data, lack a unified framework for IoT data valuation, and do not exhaustively examine how different IoT technologies interact with data valuation. Furthermore, these surveys overlook the dynamic nature of IoT environments where data is generated, processed, and utilized in real-time applications.

Our paper aims to bridge these gaps by providing a comprehensive review of current data valuation research, with a specific focus on IoT. We meticulously analyze the key and evolving roles of blockchain, wireless sensing, ML, and edge computing in IoT data valuation. Our contributions include: (1) a thorough review of existing data valuation research, with a special emphasis on its application in IoT; (2) an in-depth examination of the interplay between different IoT technologies and data valuation that provides a nuanced understanding of their roles and potential; and (3) the identification of research gaps and directions for future work, aiming to foster a deeper understanding and advance IoT data valuation. In the following sections, we present our analytical framework as shown in Fig. 1, which outlines the interconnections between IoT technologies and data valuation. By offering this comprehensive review, we aim to provide valuable references for both academic and industrial sectors, fostering the advancement of IoT data valuation and paving the way for future research.

†Xinyi Shi is a visiting student at Artificial Intelligence Research Institute and Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Shenzhen MSU-BIT University, China.

*Haihan Duan is the corresponding author (duanhaihan@smbu.edu.cn).

2770-2677/24/$31.00 ©2024 IEEE
DOI 10.1109/SmartIoT62235.2024.00090

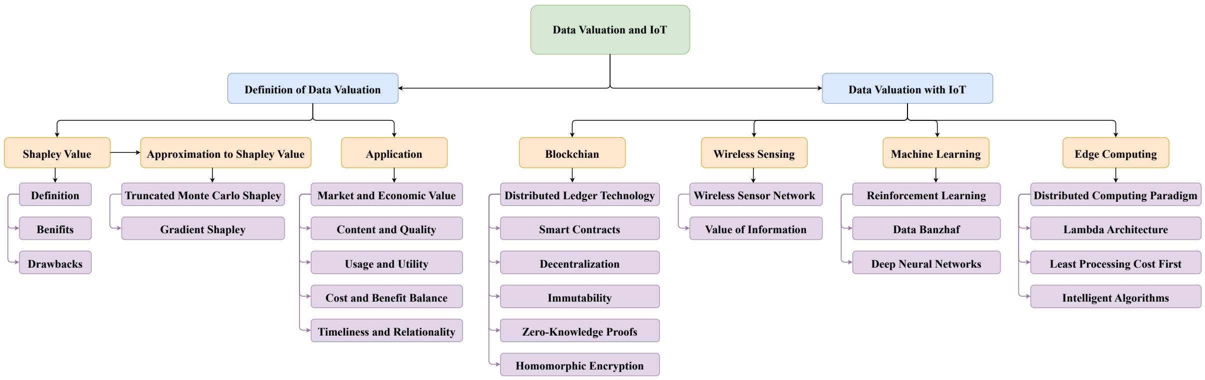

Fig. 1. Framework of Data Valuation and Prcing in IoT

II. Related Work

In recent years, the rise of large language models (LLMs) has highlighted the significance of data valuation, attracting considerable attention from the academic community. Various studies have explored aspects of data pricing and valuation across diverse domains. This paper categorizes these studies into foundational pricing strategies, valuation methods and applications, and emerging market pricing.

In 2013, Sen *et al.* [1] conducted an early survey examining several broadband pricing proposals, both static and dynamic, analyzing their implementation in consumer data plans worldwide. This foundational study provided a comprehensive overview of existing strategies, laying the groundwork for subsequent research. However, it primarily focused on consumer data plans and did not explore the complexities of data valuation in advanced data markets involving large-scale analytics or ML applications. Liang *et al.* [2] explored major concepts related to big data pricing, including digital commodity pricing principles and data market structures. Their survey encompassed economic-based and game theory-based models but did not discuss emerging technologies like blockchain, which could revolutionize data pricing through decentralized mechanisms. Gizelis *et al.* [3] conducted a comprehensive survey of pricing schemes, categorizing them into demand-based, quality-of-service (QoS)-based, resource-based, and dynamic pricing. Their analysis highlighted challenges related to fairness and technical implementation, but did not investigate the broader implications for large-scale data markets.

Ruoxi Jia *et al.* [4] investigated the Shapley Value and data valuation, focusing on suitable approximation methods from cooperative game theory. Although their research provided important insights, the computational intensity of the Shapley Value may limit its practicality in large-scale environments. Zhang *et al.* [5] introduced a model for a data market where freshness, quantified by the Age of Information (AoI), is critical. They analyzed time-dependent and quantity-based pricing schemes, demonstrating that the latter maximizes provider profit while minimizing social costs. This study offers a framework for pricing fresh data, though it primarily addresses single user-provider interactions.

Duan *et al.* [6] explored optimal investment and pricing decisions for cognitive mobile virtual network operators (C-MVNOs) under spectrum supply uncertainty. They proposed a model where C-MVNOs dynamically lease spectrum to maximize profit while serving secondary users. However, their study mainly focused on C-MVNO and secondary user interactions without discussing broader implications for large-scale data markets. Wu *et al.* [7] proposed a pricing strategy for Large Model as a Service (LMaaS), introducing the iterative model pricing (IMP) algorithm and robust selection and rental (RSR) algorithm. Their experiments demonstrated significant performance improvements, yet the study focused on provider-customer interactions without discussing implications for large-scale markets or advanced analytics integration.

Despite these contributions, existing surveys have not provided a systematic review of data valuation integration within the IoT ecosystem. Most research has focused on social applications of data valuation, leaving a gap in understanding how these concepts can be applied in IoT environments. This paper aims to bridge this gap by providing a comprehensive overview of data valuation methods tailored for IoT applications, discussing theoretical and practical aspects to enhance the utility and economic efficiency of IoT data markets.

III. Definition of Data Valuation

As data becomes the driving force behind technological advancements and economic growth, a fundamental challenge arises: quantifying the value of data in algorithmic forecasting and decision-making. However, determining a fair valuation for personal data remains elusive. Some argue that personal data should be considered individual property, warranting compensation in exchange for its use [8]. The Shapley value offers a unique payment scheme that aligns with the concept of data value, which fulfills many of its expectations. Nevertheless, calculating the Shapley value typically requires an exponential amount of time. To overcome this challenge, we apply efficient algorithms for approximating Shapley values.

A. Shapley Value

1) Definition: The Shapley value, introduced by Lloyd Shapley in 1953, investigates the challenge of fairly distributing rewards in cooperative games. The fundamental concept is to allocate rewards to each participant based on their marginal contribution to the overall output. Specifically, the Shapley value calculates the benefit due to each participant by averaging their marginal contributions across all possible sequences in which participants can join the coalition. Therefore, the Shapley's formula is given by:

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

where $\phi_i(v)$ represents the Shapley value of participant i, N denotes the set of all participants, S is a subset of the participants, $v(S)$ is the value of subset S, $|N|$ is the total number of participants, and $|S|$ is the size of subset S [9].

2) Benefits: A significant advantage of using Shapley values is that they offer more comprehensive and insightful analyses compared to popular methods such as leave-one-out scores or leverage scores when evaluating the value of data for a given learning task. For example, data points with low Shapley values can effectively capture data corruption and outliers, assisting in the identification and management of problematic data within the dataset. Conversely, data points with high Shapley values highlight the types of new data that should be acquired to enhance the predictive model. This guidance aids in directing data collection efforts, ultimately leading to improved model performance and robustness.

3) Drawbacks: However, there are notable drawbacks to using Shapley values. The primary issue is that calculating Shapley values necessitates computing all possible marginal contributions, which becomes exponentially large as the size of the dataset increases. This makes the direct computation of Shapley values highly complex and resource-intensive, posing significant challenges in practical applications.

B. The Approximation to Shapley Value

1) Truncated Monte Carlo Shapley: Truncated Monte Carlo Shapley is an approximation method for calculating Shapley values. This method leverages random sampling to efficiently estimate Shapley values, particularly for large datasets or complex models. The detailed steps are as follows:

1) **Random Sampling of Feature Subsets**: Randomly sample a subset from the set of all input features. The size of this subset can be adjusted based on computational resources and the desired accuracy.
2) **Contribution Calculation**: For each feature in the sampled subset, compute its contribution within the current subset. This can be done using the Leave-One-Out method, where the model's prediction is evaluated with and without the feature to determine its contribution.
3) **Accumulate Contributions**: Accumulate the contributions of each feature across multiple sampled subsets to obtain an approximate Shapley value for each feature.
4) **Repeat Sampling**: Repeat this process multiple times with different sampled subsets, and then average the Shapley values obtained from each sample to enhance the accuracy of the estimation [10].

2) Gradient Shapley: Gradient Shapley is a method that leverages gradient information to accelerate the computation of Shapley values, measuring the contribution of each feature to the model's prediction. The detailed steps are as follows:

1) **Model Input and Output**: Suppose we have a trained model f, an input feature vector $\mathbf{x}$, and the goal is to compute the Shapley value of each feature x_i for the model prediction $f(\mathbf{x})$.
2) **Gradient Calculation**: Calculate the gradient of the model output $f(\mathbf{x})$ with respect to each input feature x_i, denoted as $\frac{\partial f(\mathbf{x})}{\partial x_i}$. This calculated gradient represents the immediate effect of small variations in the input feature x_i on the model's overall prediction.
3) **Marginal Contribution Estimation**: For various values of each input feature x_i, perform forward and backward passes through the model to compute the corresponding gradients. This detailed gradient information is crucial for accurately estimating the marginal contributions of each feature within different feature combinations.
4) **Approximate Shapley Value**: Combine this gradient information using appropriate weighting methods to approximate the Shapley value of each feature. The core concept here involves estimating the marginal contributions through gradients rather than explicitly calculating all possible feature combinations, thereby significantly reducing the overall computational complexity.

C. Application of Data Valuation

Data valuation models offer structured approaches to assess the value of data across different fields and industries. This section explores the primary categories of these models, along with their specific applications and implications.

1) Market and Economic Value-Based Models: Market Pricing Models assess data value based on its market price, crucial for data markets and trading platforms. These models analyze market dynamics, demand, and supply to establish fair prices for data transactions. For example, pricing algorithms can determine optimal dataset prices using historical transactions and current demand, ensuring competitive pricing strategies. Furthermore, these models can incorporate external factors such as emerging trends and regulatory changes, enhancing their predictive accuracy. Economic Benefit Models evaluate the direct economic gains from data applications. For instance, businesses can enhance marketing strategies using customer data, leading to increased sales and improved customer engagement. Models like Return on Investment (ROI) quantify financial returns from data-driven initiatives, supporting enterprise decision-making and strategic planning by providing clear insights into the cost-effectiveness of data investments [11]. Additionally, understanding economic value can help organizations justify data acquisition costs and prioritize data-driven projects that yield the highest financial impact.

Ultimately, these models not only guide pricing and investment decisions but also foster a more transparent and efficient data marketplace, benefiting both buyers and sellers.

2) Content and Quality-Based Models: Content Assessment Models, such as the Laney model, evaluate data utility, business impact, and quality, considering the three Vs: volume, velocity, and variety. High-quality data improves decision-making and operational efficiency [12]. For instance, content assessment can identify gaps in a customer database, enhancing customer relationship management through targeted and personalized strategies. Moreover, these models can help organizations maintain a competitive advantage by ensuring they utilize the most relevant and high-quality data available. Information Quality Models, like the Viscusi and Batini model, assess data value based on quality, structure, and diffusion utility [13]. These models are vital for information system design and data management strategies, as they guide the selection of data sources that meet quality standards and align with organizational objectives. Ensuring that data adheres to these quality benchmarks not only bolsters analytics but also enhances the reliability of insights derived from data, leading to more informed and strategic business decisions.

3) Usage and Utility-Based Models: Usage Utility Models evaluate data value based on its contribution to specific applications, especially in data science and ML. For instance, Shapley values can identify which data points improve an ML model's performance, allowing organizations to focus on acquiring and retaining the most impactful data. Utility Maximization Models optimize the benefits derived from data, emphasizing its impact on business processes and outcomes. In healthcare, models like Data Envelopment Analysis (DEA) analyze patient data usage to improve outcomes and operational efficiency [14]. These models help organizations prioritize data that maximizes utility while also highlighting the importance of continuous evaluation to adapt to changing business needs and technology advancements. Moreover, by utilizing these models, organizations can enhance their operational frameworks, making them more resilient and responsive to market fluctuations, thus further increasing their competitive edge.

4) Cost and Benefit Balance-Based Models: Cost-Benefit Models, such as the Yanlin and Haijun framework, evaluate data value based on production costs and usage benefits [15]. These models assist organizations in determining whether the costs of maintaining large datasets are justified by the insights they yield, ensuring financial sustainability in data management practices. By analyzing both direct and indirect costs, organizations can make more informed and strategic decisions regarding data investments. Cost-Effectiveness Models emphasize balancing costs and benefits, which is particularly crucial in budget-constrained scenarios. For example, Net Present Value (NPV) models guide decisions between data storage solutions by optimizing cost-performance trade-offs [16]. This strategic approach enables organizations to allocate resources effectively, minimizing waste while maximizing the return on their data assets. Furthermore, by implementing cost-benefit analysis frameworks, organizations can identify underperforming assets and efficiently redirect investments toward more promising and impactful data-driven initiatives.

5) Timeliness and Relationality-Based Models: Timeliness Evaluation Models, like the Keisler model, assess data value based on its timeliness and usage conditions, making them ideal for time-sensitive applications like financial transactions and monitoring systems [17]. For example, the value of market data in financial trading heavily depends on its freshness, emphasizing the need for real-time data processing capabilities. Relational Evaluation Models analyze the value of data by considering interdependencies between data blocks, optimizing data architecture, and enhancing information flow efficiency [18]. By understanding these relationships, organizations can improve data integration and ensure that their data ecosystems operate cohesively, ultimately leading to better analytical outcomes and business intelligence. This approach also facilitates proactive decision-making, as organizations can respond to data trends and shifts more effectively, thereby enhancing overall operational agility and market responsiveness.

In conclusion, diverse data valuation models—including market, content, utility, cost-benefit, timeliness, and relational perspectives—play a pivotal role in optimizing data use and maximizing its value. As data becomes increasingly integral to various industries, the proper application of these models will be essential in driving both operational success and innovation.

IV. Data Valuation and Pricing with IOT

The term IoT was first introduced by Kevin Ashton in 1999 within the context of supply chain management [19]. The IoT comprises a network of interconnected devices, sensors, and systems designed to seamlessly integrate the physical and digital worlds [20]. Atzori *et al.* [21] propose that the IoT can be realized through three distinct paradigms: Internet-oriented (middleware), object-oriented (sensor), and semantic-oriented (knowledge). While the IoT represents a compelling and evolving concept, significant challenges persist in ensuring a secure, efficient, and robust ecosystem that encompasses all components of the IoT architecture [22].

A. Data Valuation with Blockchain

Blockchain technology represents a significant advancement from the distributed database technology explored since the 1970s. Distributed Ledger Technology (DLT) processes databases as distributed shared data, with blockchain being a notable example. Blockchain enhances data credibility and security by providing a transparent and immutable record system. It offers reliable proof of data provenance, ownership, and transactions, thereby reducing instances of data fraud and disputes. A smart contract is a computational process that runs upon transaction execution, involving inputs, outputs, and state changes. All blockchains incorporate smart contracts to handle transaction logic, such as verifying input signatures and matching output balances with inputs [23], [24].

The IoT faces significant challenges in establishing a secure and resilient ecosystem. Current IoT systems depend on a centralized server-client model, where devices are connected, identified, and authenticated via cloud servers, and communication occurs over the Internet. While this approach functions effectively today, it may not satisfy the future needs of expanding IoT ecosystems. Blockchain technology offers a promising solution by enabling the tracking and coordination of numerous connected devices through decentralized peer-to-peer messaging, file distribution, and autonomous operations, without relying on centralized cloud services. Additionally, blockchain facilitates smart contracts that can autonomously manage devices, reducing the need for human intervention and supporting more seamless and automated operations [22].

Furthermore, the decentralization feature of blockchain eliminates the risk of a single point of failure by recording all data operations on a distributed ledger, which make data tamper-proof. The immutability of blockchain ensures that once data is recorded, it cannot be altered, which enhances data credibility and security. In addition, blockchain's transparency allows all participants to verify and trace data flow. This prevents forgery and fraud, thereby protecting data integrity [25]. Blockchain also functions as a trading platform, using its decentralized DLT to ensure that all transaction records are permanently stored and unalterable. Each transaction undergoes verification and recording by nodes within the network, which ensures data transparency and security. Smart contracts automatically execute transaction terms, leading to increased efficiency. Because blockchain is immutable, both parties can trust the transaction data on the platform, which prevents fraud and provides a secure, transparent, and efficient trading environment [26]. Blockchain-based data exchange platforms enable decentralized data sourcing and sharing, reducing vulnerabilities associated with centralized management. Technologies like zero-knowledge proofs and homomorphic encryption provide enhanced privacy protection. Federated learning systems facilitate the sharing of algorithms without exposing raw data, further ensuring privacy [27].

In summary, blockchain technology has substantial potential to improve data security, transparency, and automated transactions, playing a crucial role in the IoT ecosystem and data valuation. By enabling decentralized control, enhancing data privacy, and ensuring fair transactions, blockchain technology paves the way for a more secure and efficient digital future.

B. Data Valuation with Wireless Sensing

Wireless sensing technology plays a crucial role in the IoT by allowing devices to transmit data without physical connections. These sensors are extensively used in environmental monitoring, health monitoring, asset tracking, and other fields. They provide a foundation for data valuation by capturing real-time data. In a wireless sensing environment, the value of data is determined by its timeliness, accuracy, and relevance.

The concept of Value of Information (VoI) in Wireless Sensor Networks (WSNs) is defined in various ways, depending on the focus area. These definitions can be categorized as follows: (1) Internal System Costs, which encompass factors such as energy consumption and network resource usage [28]; (2) Probabilistic Factors, which discuss the economic impact on decision-making processes [29]; (3) Decision-Making, which emphasizes reducing uncertainty to enhance outcomes [17]; (4) System Utility, which evaluates the usefulness of information within a specific context [30]; (5) Information Consumers, which focuses on the goals of applications and the utility for end users [31]; and (6) External System Costs, which considers factors such as information pricing and privacy costs from an economic perspective [32], [33].

Recent technological advances in low-power integrated circuits and wireless communications have made efficient, low-cost, and low-power miniature devices available for remote sensing applications. WSNs utilize these advancements to deploy numerous intelligent sensors that collect, process, and analyze data in various environments. Key components of WSNs include hardware (sensor interfaces, processing units, transceivers, power supply), communication stack (topology, routing, MAC layer, internet gateway), middleware (platform-independent service architecture for sensor applications), and secure data aggregation methods to ensure network longevity and data reliability [34]. However, WSNs still face significant resource constraints, including limited energy, bandwidth, and computational resources. To address these challenges, economic and pricing models are applied for resource optimization in several areas: resource allocation, which adapts to variable resource availability using pricing mechanisms; energy control, which uses utility-based pricing strategies to optimize sensor power levels; and task allocation, which aims to balance energy consumption and minimize delays by dynamically adjusting to resource changes [35].

Wireless sensing technology enhances IoT by enabling real-time data transmission, which is vital for applications like environmental and health monitoring. VoI in WSNs is evaluated through various models that investigate system costs, decision-making processes, and user utility.

C. Data Valuation with Machine Learning

With the large amount of data generated by IoT devices, ML has become an important tool for data analysis and valuation. ML algorithms can extract valuable information and patterns from vast amounts of data, thereby enhancing its value. In smart cities, ML algorithms can optimize traffic flow, predict environmental changes, and improve urban management efficiency. In industrial IoT, ML can perform predictive maintenance by analyzing equipment data, thereby reducing equipment failures and downtime. Moreover, ML can determine dynamic pricing of data by analyzing market demand and data quality in real time, which enables the development of reasonable pricing strategies for data transactions. This data-driven approach not only increases data utilization but also enhances the vitality of the data market.

Key components of data valuation—namely quality, quantity, diversity, and freshness—have a substantial impact on ML. Data quality refers to the accuracy, completeness, and consistency of data, directly affecting model training effectiveness. Data quantity relates to the amount of data; more data generally enhances model performance, though redundancy must be considered. Data diversity pertains to the variety of features and distributions within the data, helping improve model generalization. Data freshness is crucial for certain applications, enabling models to stay updated [36].

Yoon *et al.* [37] designed Data Valuation using Reinforcement Learning (DVRL), a novel method that uses reinforcement learning to evaluate data value. The agent selects data points for model training and adjusts its strategy based on the model's performance on a validation set, learning to identify high-value data points. Experimental results show that DVRL outperforms traditional methods in data selection and improving model overall performance. The Banzhaf value is a metric used to measure the influence of players in cooperative games. It evaluates the importance of each player by calculating how often they are pivotal in all possible coalitions. Specifically, the Banzhaf value measures how many times a player can change the outcome of a decision, thus determining their relative contribution to the overall decision-making process. This method helps identify which players have greater influence and decision-making say in the decision-making process [38]. Data Banzhaf is a new framework that uses the Banzhaf value from game theory to evaluate data value. This method treats data points as participants in a cooperative game, calculating each point's Banzhaf value to measure its contribution to model performance, thereby offering a more robust and fair data valuation mechanism [39]. Furthermore, Davinz is a novel method for data valuation at the initialization stage of deep neural networks. By analyzing the network's initial response to data points and using gradient information, it assesses the value of data, thereby reducing computational complexity and enhancing the efficiency of data selection [40].

In conclusion, data valuation in ML is a complex and evolving field that integrates diverse methods to assess data significance. Approaches such as performance-based, information-theoretic, game-theoretic, and reinforcement learning provide robust frameworks for evaluating data value. As IoT generates vast amounts of data, effective valuation strategies will be essential for maximizing data utility, optimizing ML models, and advancing various domains. The continued development of these strategies will enhance the accuracy, fairness, and efficiency of data-driven decision-making, supporting the broader adoption and success of ML applications.

D. Data Valuation with Edge Computing

As the number of services increases, critical needs in IoT architecture include reducing data latency and improving resource utilization efficiency. Edge computing's three-tier architecture includes terminal, edge, and cloud layers. The terminal layer collects data and sends it to the edge layer for processing and storage, while the cloud layer handles large-scale centralized processing [41]. Edge computing is a distributed computing paradigm that shifts data processing and analysis from central data centers to the network edge, closer to where the data is generated. This approach is particularly advantageous in IoT environments, as it reduces data transmission latency, enhances real-time processing capabilities, and efficiently manages large data volumes. By placing substantial computing and storage resources—referred to as clouds, microdata centers, or fog nodes—at the edge of the Internet, near mobile devices or sensors, edge computing optimizes performance. Cloudlets, small data centers located at the network edge, provide low-latency and high-bandwidth computing resources in close proximity to data sources and user devices. They act as an intermediary layer between edge devices and cloud data centers, facilitating real-time data processing and localized services. This is essential for applications such as supply chain tracking, point of sale systems, augmented reality, autonomous driving, smart cities, and distributed artificial intelligence processing [42] [43].

The Lambda architecture is a method to address pricing issues in large-scale data processing. This architecture consists of the batch layer, speed layer, and serving layer. The batch layer processes historical data using distributed file systems and batch processing frameworks, generating precomputed views. The speed layer handles real-time data using stream processing frameworks, creating real-time views. The serving layer merges data from both the batch and speed layers to provide a unified query interface for users. This Lambda architecture enables efficient, accurate, and real-time data analysis to support complex pricing strategies [44]. Least Processing Cost First (LPCF) is an optimization algorithm designed to balance energy consumption, processing capability, and data transport efficiency in edge-fog computing environments. The algorithm dynamically adjusts task allocation strategies by analyzing task characteristics (such as computational complexity and data volume) and the resource status of each node, reducing energy consumption, improving processing efficiency, and minimizing data transmission delays compared to traditional static task allocation methods [45]. Dautov *et al.* [46] propose an automated task allocation method focused on optimizing IoT data-intensive applications in clustered edge computing environments. The method employs intelligent algorithms (such as genetic algorithms and heuristic methods) to dynamically assign tasks to edge computing nodes, taking into account task resource requirements, node computing power, and network latency. This approach aims to enhance overall system resource utilization, reduce task completion time, and improve application response times.

The domain of IoT architecture and edge computing is both complex and rapidly advancing, offering diverse methods to tackle challenges in data processing and resource management. Techniques such as the three-tier architecture of edge computing, the Lambda architecture for large-scale data handling, and sophisticated task allocation algorithms collectively provide robust solutions for enhancing performance and minimizing latency. As IoT systems generate increasingly large volumes of data, these methodologies will be vital for optimizing real-time processing, improving resource utilization, and managing intricate data tasks. The continuous development and refinement of these approaches will facilitate more efficient and scalable IoT systems, thereby supporting the effective deployment and growth of applications across various fields.

V. Vision of Data Valuation

With advancements in LLMs, network capacity, and computational power, data valuation is set for even more significant improvements across various sectors beyond finance. The complexity of data valuation arises from the diversity and heterogeneity of data sources. Data is generated from numerous platforms, devices, and sources, each with unique formats, quality, and reliability. Integrating these disparate datasets to derive meaningful valuation metrics presents substantial challenges, particularly with the rapid growth of unstructured data, such as social media posts and IoT sensor data. Traditional methods often struggle to fully capture the value embedded in these diverse data types, hindering accurate and comprehensive valuation strategies. The following subsections outline three key directions for enhancing data valuation.

A. Privacy and Compliance

Different countries and regions enforce varying laws and regulations on data privacy and protection, such as the General Data Protection Regulation (GDPR), creating significant challenges for data collection, processing, and valuation. Ethical considerations also arise, particularly regarding fair data use, sharing practices, and personal privacy protection. Thus, privacy and compliance are critical aspects of data valuation. Organizations must establish robust data governance mechanisms to ensure compliance with relevant regulations and secure data handling practices [47]. This process can be resource-intensive and may necessitate significant overhauls of existing data management practices, which can strain organizational resources. Furthermore, organizations must navigate challenges associated with cross-border data transfers and processing. In a globalized business environment, the need for free data flow is evident, but balancing this need with privacy and compliance remains complex. Continuous monitoring of evolving data privacy laws in different jurisdictions is essential. Organizations should invest in developing flexible data management systems that can swiftly adapt to changing compliance requirements while maintaining effective operational capabilities.

B. Dynamic and Uncertain Nature

The value of data is not static; it evolves over time due to market demands and technological advancements, complicating accurate and timely assessment. Throughout the data lifecycle—from generation and storage to utilization—data value fluctuates, necessitating frequent and continuous evaluation and management. To address this dynamic nature, organizations must develop flexible data valuation models that can adapt to various temporal and situational contexts [48]. This adaptability often requires significant investment in research and development, making it a complex undertaking. Additionally, market demands and emerging technologies can drastically influence data value. For instance, advancements in artificial intelligence can transform previously low-value data into highly valuable assets, while traditionally high-value data may depreciate due to shifts in consumer behavior or market trends. Organizations must actively monitor market dynamics and technological changes, adjusting their data valuation strategies accordingly to ensure relevance and accuracy. The challenge lies in effectively integrating real-time market insights into existing valuation frameworks, which requires agility and responsiveness in decision-making processes.

C. Lack of Tools and Standards

The current lack of uniform tools and standards for data valuation results in inconsistent methodologies across different enterprises, leading to poor comparability and reliability of valuation outcomes. This inconsistency undermines the transparency and fairness of data transactions, which in turn stifles the growth of the data market. To address this issue, establishing standardized data valuation methods and tools is essential. Industry stakeholders must collaborate to develop universally accepted standards for data classification, valuation models, and data quality assessments. However, achieving consensus on such standards can be challenging due to the diverse interests of stakeholders and the rapid pace of technological evolution. Developers and researchers should focus on creating efficient and accurate data valuation tools that enable organizations to assess data more effectively. Moreover, industry associations and standardization bodies play a crucial role in promoting the establishment and dissemination of data valuation standards. By fostering collaboration among various entities, these organizations can help ensure that the established standards meet the diverse needs of the industry while facilitating the healthy development of the data market. By enhancing transparency and consistency in data transactions through standardized tools and methods, organizations can strengthen trust in the data market, thereby promoting the circulation and utilization of data to maximize its overall value.

VI. Conclusion

Data valuation is crucial in modern business, influencing strategic decision-making, investment, market pricing, and risk management. It has become essential in the IoT ecosystem and will increasingly optimize and monetize IoT applications. However, it faces challenges, including complex valuation methods, diverse legal and ethical issues, the dynamic nature of data value, and technological limitations. This paper reviews existing data valuation methodologies and frameworks, exploring their integration with IoT systems. It also presents our vision for future development in the IoT domain, emphasizing dynamic and adaptive data valuation, privacy and compliance, and widely standardized valuation tools. We believe the growing IoT industry will benefit from accurate, real-time data valuation, as evolving data forms, values, and meanings will ultimately enhance efficiency, security, and economic value, driving continuous innovation across various sectors.

Acknowledgment

This work was supported by Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University.

References

[1] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "A survey of smart data pricing: Past proposals, current plans, and future trends," *Acm computing surveys (csur)*, vol. 46, no. 2, pp. 1–37, 2013.

[2] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, and W. Zhao, "A survey on big data market: Pricing, trading and protection," *Ieee Access*, vol. 6, pp. 15 132–15 154, 2018.

[3] C. A. Gizelis and D. D. Vergados, "A survey of pricing schemes in wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 1, pp. 126–145, 2010.

[4] R. Jia, D. Dao, B. Wang, F. A. Hubis, N. Hynes, N. M. Gürel, B. Li, C. Zhang, D. Song, and C. J. Spanos, "Towards efficient data valuation based on the shapley value," in *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 1167–1176.

[5] M. Zhang, A. Arafa, J. Huang, and H. V. Poor, "How to price fresh data," in *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. IEEE, 2019, pp. 1–8.

[6] L. Duan, J. Huang, and B. Shou, "Investment and pricing with spectrum uncertainty: A cognitive operator's perspective," *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1590–1604, 2011.

[7] P. Wu, Q. Liu, Y. Dong, Z. Wang, and F. Wang, "Lmaas: Exploring pricing strategy of large model as a service for communication," *IEEE Transactions on Mobile Computing*, 2024.

[8] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, vol. 10, no. 3152676, pp. 10–5555, 2017.

[9] M. Sundararajan and A. Najmi, "The many shapley values for model explanation," in *International conference on machine learning*. PMLR, 2020, pp. 9269–9278.

[10] A. Ghorbani and J. Zou, "Data shapley: Equitable valuation of data for machine learning," in *International conference on machine learning*. PMLR, 2019, pp. 2242–2251.

[11] J. Boyd, R. Epanchin-Niell, and J. Siikamäki, "Conservation planning: a review of return on investment analysis," *Review of Environmental Economics and Policy*, 2015.

[12] D. B. Laney, *Infonomics: how to monetize, manage, and measure information as an asset for competitive advantage*. Routledge, 2017.

[13] G. Viscusi and C. Batini, "Digital information asset evaluation: Characteristics and dimensions," in *Smart Organizations and Smart Artifacts: Fostering Interaction Between People, Technologies and Processes*. Springer, 2014, pp. 77–86.

[14] S. Kohl, J. Schoenfelder, A. Fügener, and J. O. Brunner, "The use of data envelopment analysis (dea) in healthcare with a focus on hospitals," *Health care management science*, vol. 22, pp. 245–286, 2019.

[15] W. Yanlin and Z. Haijun, "Data asset value assessment literature review and prospect," in *Journal of Physics: Conference Series*, vol. 1550, no. 3. IOP Publishing, 2020, p. 032133.

[16] C. A. Magni, "Investment decisions, net present value and bounded rationality," *Quantitative Finance*, vol. 9, no. 8, pp. 967–979, 2009.

[17] J. M. Keisler, Z. A. Collier, E. Chu, N. Sinatra, and I. Linkov, "Value of information analysis: the state of application," *Environment Systems and Decisions*, vol. 34, pp. 3–23, 2014.

[18] M. Bendechache, J. Attard, M. Ebiele, and R. Brennan, "A systematic survey of data value: Models, metrics, applications and research challenges," *IEEE Access*, 2023.

[19] K. Ashton *et al.*, "That 'internet of things' thing," *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009.

[20] S. Li, L. D. Xu, and S. Zhao, "The internet of things: a survey," *Information systems frontiers*, vol. 17, pp. 243–259, 2015.

[21] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[22] E. Karafiloski and A. Mishev, "Blockchain solutions for big data challenges: A literature review," in *IEEE EUROCON 2017-17th International Conference on Smart Technologies*. IEEE, 2017, pp. 763–768.

[23] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi, and J. Wang, "Untangling blockchain: A data processing view of blockchain systems," *IEEE transactions on knowledge and data engineering*, vol. 30, no. 7, pp. 1366–1385, 2018.

[24] W. Cai, Z. Wang, J. B. Ernst, Z. Hong, C. Feng, and V. C. Leung, "Decentralized applications: The blockchain-empowered software system," *IEEE access*, vol. 6, pp. 53 019–53 033, 2018.

[25] K. Wang, J. Dong, Y. Wang, and H. Yin, "Securing data with blockchain and ai," *Ieee Access*, vol. 7, pp. 77 981–77 989, 2019.

[26] S. Casale-Brunet, M. Mattavelli, and L. Chiariglione, "Exploring blockchain-based metaverses: Data collection and valuation of virtual lands using machine learning techniques," *Digital Business*, vol. 3, no. 2, p. 100068, 2023.

[27] W. M. Charles and B. M. Delgado, "Health datasets as assets: Blockchain-based valuation and transaction methods," *Blockchain in Healthcare Today*, vol. 5, 2022.

[28] A. Mahajan, "Remote estimation over control area networks," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2017, pp. 1–5.

[29] R. A. Howard, "Information value theory," *IEEE Transactions on systems science and cybernetics*, vol. 2, no. 1, pp. 22–26, 1966.

[30] S. Bharti, K. K. Pattanaik, and P. Bellavista, "Value of information based sensor ranking for efficient sensor service allocation in service oriented wireless sensor networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 823–838, 2019.

[31] L. Bölöni, D. Turgut, S. Basagni, and C. Petrioli, "Scheduling data transmissions of underwater sensor nodes for maximizing value of information," in *2013 IEEE global communications conference (GLOBECOM)*. IEEE, 2013, pp. 438–443.

[32] S. Tan, C. Jiang, and Y. He, "A research on information value in energy internet based on risk theory," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*. IEEE, 2017, pp. 1–6.

[33] F. Alawad and F. A. Kraemer, "Value of information in wireless sensor network applications and the iot: A review," *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9228–9245, 2022.

[34] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[35] N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Data collection and wireless communication in internet of things (iot) using economic analysis and pricing models: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2546–2590, 2016.

[36] R. H. L. Sim, X. Xu, and B. K. H. Low, "Data valuation in machine learning:" ingredients", strategies, and open challenges." in *IJCAI*, 2022, pp. 5607–5614.

[37] J. Yoon, S. Arik, and T. Pfister, "Data valuation using reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2020, pp. 10 842–10 851.

[38] E. Lehrer, "An axiomatization of the banzhaf value," *International Journal of Game Theory*, vol. 17, pp. 89–99, 1988.

[39] J. T. Wang and R. Jia, "Data banzhaf: A robust data valuation framework for machine learning," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023, pp. 6388–6421.

[40] Z. Wu, Y. Shu, and B. K. H. Low, "Davinz: Data valuation using deep neural networks at initialization," in *International Conference on Machine Learning*. PMLR, 2022, pp. 24 150–24 176.

[41] G. Kaur and R. S. Batth, "Edge computing: Classification, applications, and challenges," in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*. IEEE, 2021, pp. 254–259.

[42] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang, "A survey on the edge computing for the internet of things," *IEEE access*, vol. 6, pp. 6900–6919, 2017.

[43] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.

[44] M. Femminella, M. Pergolesi, and G. Reali, "Iot, big data, and cloud computing value chain: pricing issues and solutions," *Annals of Telecommunications*, vol. 73, pp. 511–520, 2018.

[45] N. Mohan and J. Kangasharju, "Placing it right!: optimizing energy, processing, and transport in edge-fog clouds," *Annals of Telecommunications*, vol. 73, pp. 463–474, 2018.

[46] R. Dautov and S. Distefano, "Automating iot data-intensive application allocation in clustered edge computing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 1, pp. 55–69, 2019.

[47] A. S. Alenizi and K. A. Al-Karawi, "Internet of things (iot) adoption: challenges and barriers," in *Proceedings of Seventh International Congress on Information and Communication Technology: ICICT 2022, London, Volume 3*. Springer, 2022, pp. 217–229.

[48] H. He, K. Xu, and Y. Liu, "Internet resource pricing models, mechanisms, and methods," *Networking Science*, vol. 1, pp. 48–66, 2012.