

A BRIEF SURVEY OF DATA PRICING FOR MACHINE LEARNING

Zuoqi Tang¹, Zheqi Lv², Chao Wu^{3,4*}

¹Department of Computer Science and Technology, Zhejiang University, China

²Department of Marine Informatics, Zhejiang University, China

³Department of Public Affairs, Zhejiang University, China

⁴Center of Social Welfare and Governance, Zhejiang University, China

ABSTRACT

Big data and machine learning are poised to revolutionize the field of artificial intelligence and represent a step towards building an intelligent society. Big data is considered to be the key to unlocking the next great waves of growth in productivity, the value of data is realized through machine learning.

In this survey, we begin with an introduction to the general field of data pricing and distributed machine learning then progress to the main streams of data pricing and mechanism design methods. Our survey will cover several current areas of research within the field of data pricing, including the incentive mechanism design for federated learning, reinforcement learning, auction, crowdsourcing, and blockchain, especially, focus on reward function for machine learning and payment scheme. In parallel, we highlight the pricing scheme in data transactions, focusing on data evaluation via distributed machine learning. To conclude, we discuss some research challenges and future directions of data pricing for machine learning.

KEYWORDS

Data pricing, Big data, Machine learning, Data transaction

1. INTRODUCTION

Nowadays, we live in the era of mobile Internet, big data and artificial intelligence, these technologies are changing our lives.

In the era of the mobile Internet, especially the coming 5G and the Internet of Things, which is more open and more connected. In this scenario, the data which distributed the different edging nodes is exploding exponentially, these data belong to the different organizations and individuals. Data collection and data value become more important for economic and social activities.

According to the report of the Forbes, the worldwide Big Data market revenues for software and services are projected to increase from \$42B (Billion) in 2018 to \$103B in 2027, attaining a Compound Annual Growth Rate (CAGR) of 10.48%. As a part of this forecast, Wikibon estimates the Worldwide Big Data market is growing at an 11.4% CAGR between 2017 and 2027, growing from \$35B to \$103B, as illustrated in figure 1. At the same time, artificial intelligence (AI) is transforming economies and societies, changing the way we communicated and work. As the machine learning models are growing larger and more complex, it requires a variety of data, which is heterogeneous. How to effectively process data while protecting privacy is a huge challenge in the field of machine learning for academic and industrial circles.

Natarajan Meghanathan et al. (Eds): SIPP, BIGML, DaKM, SOEN, AISC - 2020

pp. 99-110, 2020. © CS & IT-CSCP 2020

DOI: 10.5121/csit.2020.100209

So, in the age of data-driven technologies, data pricing or incentive is an important topic between data consumers and data owners for machine learning. One of the central question in machine learning deals with pricing or reward the behavior of a data provider. In addition to the perspective of the AI, data pricing would be key to support the setup as well as to sustain a good ecosystem for data markets.

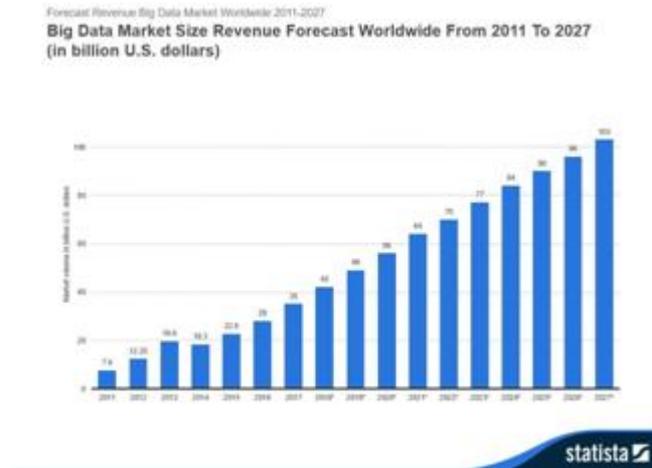


Figure 1. Revenue forecast (Source: Wikibon and reported by Statista)

The remainder of this survey is organized as follows: In section 2, we briefly introduce the basic concept related to the data pricing for machine learning and the research progress of data pricing. In section 3, we review the mechanism design in artificial intelligence. In section 4, we review the pricing mechanism in the data transaction. In section 5, we discuss the challenges and future directions. Finally, we conclude the paper in section 6.

2. DATA PRICING

In this section, we review some research progress of data pricing.

2.1. Data Market

The amount of collected data in our world has been exploding due to a number of new applications, especially mobile applications and the Internet of Thing-based smart systems. With the exponential growth of data, how to efficiently utilities the data becomes a critical issue. This calls for the development of a big data platform that enables efficient data utilization between data providers and data consumers (Machine learning) [15]. In the data market, three agents are involved, namely, the seller who provides the datasets, the buyer who is interested in buying ML models instance, and the market who interacts between the seller and buyer. In data markets, how to verify whether the service provider has truthfully collected and processed data becomes a pressing problem for the data consumer. Niu et al. [20] proposed the first secure mechanism TPDM (integrating Truthfulness and Privacy preservation in Data Markets) for personal data markets, achieving both truthfulness and privacy preservation.

The Internet of Things (IoT) has emerged as a new paradigm for the future Internet. In IoT, devices are connected to the Internet and thus are a huge data source for numerous applications. In [22], Niyato et al. focused on addressing data management in IoT through using a smart data pricing approach and proposed a new pricing scheme for IoT service providers to determine the sensing data buying price and IoT service subscription fee offered to sensor owners and service user.

In [21], the authors introduced a big data market model, which is composed of a data source, a service provider, and consumer, and proposed utility functions of data when the data is used in big data analytics. Based on the data utility functions, the authors developed an optimal pricing scheme that allows the service provider to determine the amount of data to be acquired to provide services to users. Additionally, the authors showed the Stackelberg game can be model the strategy of the data source to achieve the maximum profit.

2.2. Categories of the Data Pricing Models

There are two pricing models according to the different principles. The traditional models are economic-based pricing, which establishes the price strategies based on classic economic theory. The second model is game theory-based pricing, which is the dynamic price in competitive scenarios [17].

The game theory finds nowadays a broad range of applications in engineering and machine learning. The Game theory being a mathematical tool to analyze strategic interactions between rational decision-makers, in this survey, we review the usage of Game Theory in different machine learning settings involving usage of large amounts of data[23]. The goal is to provide an overview of the use of game theory in different applications that rely extensively on big data. In the traditional data markets, Muschalle et al. [18] listed the six main categories of pricing models.

- **Free** data can be obtained from public authorities.
- **Usage-Based prices** correspond to the human rationality that every single unit of a commodity raises the total amount of money to pay for.
- **Package pricing** refers to a pricing model that offers a customer a certain amount of data for a fixed fee.
- **Flat fee tariff** is one of the simplest pricing models with minimal transaction costs. It is based on time as the only parameter.
- **Two-Part Tariff** is a combination of package pricing and flat pricing strategies.
- **Freemium** is another approach of pricing data and algorithms on marketplaces. The idea is to let users join and use basic services for free and charge them for premium services that provide additional value to customers.

On the other hand, Raskar et al. [24] proposed a thorough data pricing strategy that needs to adhere to the following guiding principles.

- **Liquidity**: models freshness of data in terms of value vs diminished/increase value over time.
- **Traceability**: can be only 'sold' once, or sold non-exclusively.
- **Consent**: maintains the privacy of the owner, tracks consent over time, and reduces.
- **Neutrality**: accessible to all buyers to prevent unfair trading practices.
- **Resource**: allows for calling back, providers right to be forgotten, allows for some course correction, broadly remains self-sustaining.

2.3. Data Pricing for Private Data

Personal data has value to both its owner and to institutions. Li et al. [13] introduced a framework for selling private data. Buyers can purchase any liner query, with any amount of perturbation, and need to pay accordingly. Data owners, in turn, are compensated according to the privacy loss they incur for each query. Moreover, with the help of comparative analysis of existing data pricing models and strategies, Shen et al. [29] proposed a pricing model for Big Data based on tuple granularity, which includes information entropy, weight value, data reference index, cost, and credit rating.

3. MECHANISM DESIGN IN AI

Mechanism design is a field in economics and game theory that takes an engineering approach to design economic mechanisms or incentives, toward desired objectives, in strategic settings, where players act rationally. Because it starts at the end of the game, then goes backward, it is also called reverse game theory. Mechanism design studies solution concepts for a class of private-information games.

In this section, we review state-of-the-art research works on the application of the Mechanism design.

3.1. Incentive Mechanism in Federated Learning

Today's artificial intelligence still faces two major challenges. One is that, in most industries, data exists in the form of isolated islands. The other is how to price data value in the different organizations. Most existing data pricing methods are very ineffective and unavailable for training Machine learning models when we cannot directly access the training dataset. The Federated Learning (FL) [11] first proposed by Google in 2016, which is a possible solution to these challenges. Next, we will make a brief introduction to distributed machine learning and federated learning.

Distributed machine learning commonly refers to multi-node machine learning algorithms and architecture, which are designed to improve performance, increase accuracy, and adapt to bigger train dataset or models. The architecture of the distributed machine learning is shown in Fig 2.

- Data/Model aggregation
- Single machine/model optimization
- Submodel/Local Data
- Data/Model partition

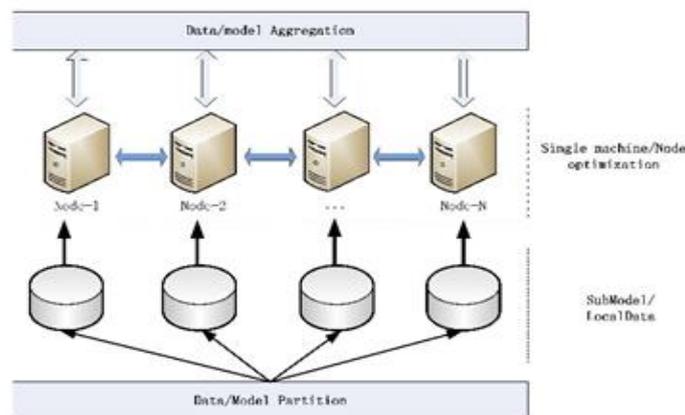


Figure 2. Distributed machine learning architecture

Federated Learning is a distributed machine learning approach that enables model training on a large corpus of decentralized data [2]. Federated Machine learning (FML) [34] creates an ecosystem for multiple parties to collaborate on the building models while protecting data privacy for the participants. Instead of transferring data directly into a centralized data warehouse for building machine learning models, Federated Learning allows each party to own the data in its place and still enables all parties to build a machine learning model together. This method stands in contrast to the traditional centralized machine learning approach where all data samples are uploaded to one data center.

To fully commercialize federated learning among different organizations, a fair platform, and incentive mechanism needs to be developed. After the model simulations built, the performance of the model will be manifested in the actual application. This performance can be recorded in a permanent data recording mechanism. Organizations that provide more data will be better off, and the model's effectiveness depends on the data provider's contribution to the system. The effectiveness of these models is distributed to parties based on the federated mechanism and continue to motivate more organizations to join the data federation [37].

Currently, federated learning is an emerging machine learning technique, most of the existing studies mainly focus on optimizing federated learning algorithms to improve model training performance, data privacy, and security. However, incentive mechanisms to motivate mobile edging smart devices to join model training have been largely overlooked. Edging devices suffer from considerable overhead in terms of computation and communication during the federated mode learning process. Without a well-designed incentive, self-interested mobile devices will be unwilling to join federated learning tasks. In [10], Kang et al. adopted the contract theory to design an effective incentive mechanism for simulating mobile devices with high-quality data participate in federated learning.

3.2. Incentive Mechanism in Reinforcement Learning

The design of reward functions in reinforcement learning (RL) is a human skill that comes with experience. Unfortunately, there is not any methodology in the literature that could guide a human to design the reward function. In [4], Clayton et al. used systematic instructional Design, an approach in human education, to engineer a machine education methodology to design reward functions for reinforcement learning. The methodology can guide the design of hierarchy learning incrementally through a multi-part reward function. The hierarchy acts as a decision fusion function that combines the individual behaviors and skill learn by each instruction to create a smart shepherd to control the swarm.

3.3. Mechanism Design in Auction

Auctions are protocols to allocate goods to buyers who have preferences over them and collect payments in return. Economists have invested significant effort in designing auction rules that result in allocations of the goods that are desirable for the group as a whole. Mechanism design is a field in economics that deals with setting incentives and interaction rules among self-interested agent samples to achieve desired objects for the group as a whole. In [33], Teerapittayanon et al. proposed a deep learning-based approach to automatically design auctions in a wide variety of domains, shifting the design work from human to machine. On the other hand, considering the traditional markets, it is extremely difficult for e-commerce companies to adjust prices when they receive more information from consumers. Shen et al. [28] proposed a reinforcement mechanism design [32] to tackle the dynamic pricing problem in sponsored Search auctions.

3.4. Mechanism Design in Crowdsourcing

An interesting recent scenario of inaccurate supervision occurs with crowdsourcing, a popular paradigm to outsource work to the individual. For machine learning, crowdsourcing is commonly used as a cost-saving way to collect labels for training data. The ability to quickly collect large-scale and high labeled datasets is crucial for Machine Learning. In online crowdsourcing, designing optimal pricing policies and determining the right monetary incentives is central to maximizing the requester's utility and workers' profit. To address these questions, regret minimization mechanisms are presented, which combine procurement auctions and multi-armed

bandits [30]. Considering existing mechanisms are often developed as a one-shot static solution, assuming a certain level of knowledge about worker models. In [6], a reinforcement incentive learning (RIL) method was proposed, to uncover how workers respond to different payments. RIL dynamically determines the payment without accessing any ground truth labels, and RIL can incentivize rational workers to provide high-quality labeled. Designing an effective Crowdsourcing protocol is important. In [26], a ‘double or nothing’ incentive-compatible mechanism is proposed to ensure workers behave honestly based on their self-confidence; this protocol is provable to avoid spammers from the crowd, under the assumption that every worker wants to maximize their expected payment. In [40, 39, 41, 42, 44, 43], Zheng et al., the authors leverage the tools of game theory and mechanism design to analyze the interaction of rational and selfish mobile users, then design efficient incentive mechanisms for four classical and representative applications in mobile Internet: dynamic spectrum redistribution, mobile crowdsensing, data marketplace, and cloud bandwidth management, to stimulate selfish mobile users to cooperate, achieving a win-win situation.

With the rapid growth of smart IoT devices, a mobile crowd-sensing is becoming an important paradigm to acquire information from the physical environment. However, it is challenging to estimate the data quality without the availability of ground truth data. Liu et al. [16] proposed a context-aware data quality estimation in an online manner. Zheng et al. [43] presented the first architecture of the mobile crowd-sensed data market, and conduct an in-depth study of the design problem of online data pricing. A novel online query-based crowd-sensed data pricing mechanism was proposed to determine the trading price of crowd-sensed data.

To improve the efficiency and utility of mobile Crowdsourcing system, Wang et al. [35] proposed an incentive mechanism that selects the worker candidates statically, and then dynamically selects winners after bidding. The proposed incentive mechanism includes two algorithms which are an improved two-stage auction algorithm and a truthful online reputation updating algorithm.

3.5. Reward Mechanism in Blockchain

In traditional blockchain [19], proof-of-work (PoW) incentivizes people to participate in the consensus protocol for a reward. Teerapittayanon et al. [33] introduced DaiMoN, a decentralized artificial intelligence model Network, which incentivizes peer collaboration in improving the accuracy of machine learning models. It is an autonomous Network where peers may submit models with improved accuracy and other peers may verify the accuracy improvement. DaiMoN rewards these contributing peers with cryptographic tokens by using a novel learnable Distance Embedding for Labels (DFL) function.

As the core issue of blockchain, mining requires solving a proof-of-work puzzle, which is resource expensive to implement in mobile devices due to the high computing power needed. Thus, the development of blockchain in the mobile application is restricted. To support offloading from mobile blockchain mining, an optimal pricing-based edge computing resource management approach was proposed. In [36], a two-stage Stackelberg game was adopted to jointly maximize the profit of Edge computing service Provider (ESP) and the individual utilities of different miners.

4. PRICING MECHANISM IN DATA TRANSACTIONS

This section reviews some research progress of data pricing, focusing on two types of data pricing: data-based pricing and model-based pricing.

4.1. Data-based Pricing

Most online markets are characterized by competitive settings and limited demand information. Due to the complexity of such markets, efficient pricing strategies are hard to derive. Schlosser et al. [25] analyzed stochastic dynamic pricing models in competitive markets with multiple offer dimensions, such as price, quality, and rating. Yu et al. [38] presented a bi-level mathematical programming model for the data-pricing problem that considers both data quality and data versioning strategies and a genetic algorithm was used to solve the model.

4.2. Machine-learning-based Pricing

Currently, there are two main solutions to train the machine learning model in the literature. The first is the traditional machine learning modeling method, which collects local data and uploads to the data center and then trains the machine learning model, namely centralized machine learning. The second is the distributed machine learning method, as known as decentralized machine learning. The data is at the edge node. The submodel is trained at the edge nodes and then the submodel is assembled in the central server. Moreover, collecting data through the data market and crowdsourcing is an effective method for centralized machine learning, however, decentralized machine learning is realized through blockchain or federated learning. The methods of machine learning modeling are shown in figure 3. Regardless of the way, how to do data incentives or pricing is the core and key issue for solving machine learning.

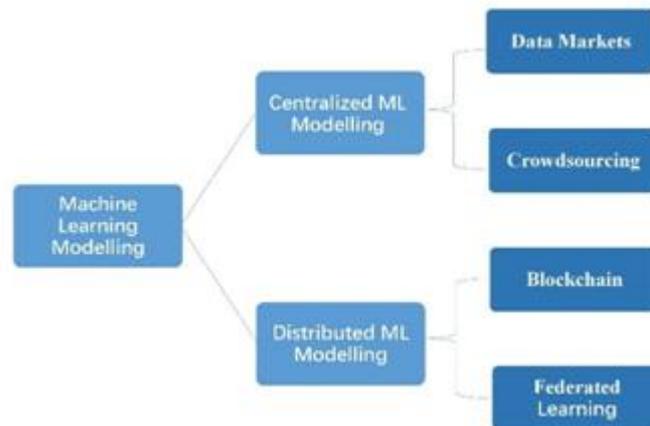


Figure 3. Methods of machine learning modeling

Data analytics using machine learning (ML) has become ubiquitous in science and engineering domain. While a lot of work focuses on reducing the training cost and storage cost of ML models, little work studies how to reduce the cost of data acquisition. Especially in the age of the Internet, data is increasingly concentrated in large firms; in the era of the Internet of things, tremendous sensor data distributed in the edging smart devices. For startups, small organizations, and institutes it is increasingly difficult to compete as the lack of availability of data can stymie any and all efforts to build a better machine-learning algorithm. Algorithm capability indeed increases with the availability and quality of the data. One way to tackle this is the marketplace approach. By creating conditions such that data, the raw material for Artificial Intelligence, can be bought and sold with security, privacy, and consent safeguard. Koutris et al. [12] proposed a framework for pricing data on the Internet that, given the price of a few views, allows the prices of any query to be derived automatically, and proposed a generalized chain queries and obtain the price by computing the time complexity of the query. Chen et al. [3] proposed a model-based pricing (MBP) framework, which instead of pricing the data, directly prices model instance. The price should depend on the accuracy of the model purchased, and not the underlying datasets.

The Shapley value [27], which is originated from coalitional game theory with proven theoretical properties, provides an effective approach to distribute contribution among features in a fair way by assigning to each feature a number which denotes its influence.

In data Markets, a fundamental challenge is how to quantify the value of data in algorithm predictions and decisions. Ghorbani et al. [5] developed a principle framework to address data valuation in the context of Supervised machine learning. Given a learning algorithm trained on n data points to produce a predictor, data Shapley was proposed to quantify the value of each training datum to the predictor performance. Although, cooperative game theory suggests as a unique way to distribute payment to data contributors such that some important theoretical properties are satisfied-data Shapley value uniquely satisfies several natural properties of equitable data valuation. Unfortunately, computing Shapley Value exactly is prohibitively expensive, therefore authors developed a sampling-based approximation algorithm-Truncated Monte Carlo Shapley. Besides, "How much is my data worth?" is an increasingly common question posed by organizations and individuals. To solve this question of fairly distributing profits among multiple data contributors, Jia et al. [9] studied the problem of data valuation by utilizing the Shapley Value, a popular notion of value that originated in cooperative game theory. In [8], jia et al. categorized firstly a data valuation problem according to whether data contributors are valued in tandem with a data analyst; where each data contributor provides a single data instance or multiple ones; whether the underlying ML model is a weighted KNN or unweighted; and whether the model solves a regression or a classification task. Then introduced two game-theoretical models for distributing the gains from an ML model and would like to understand how the shares of the analyst and the data contributors differ in the two models.

5. CHALLENGES AND FUTURE DIRECTIONS

This section discusses some research challenges and future directions of data pricing, focusing on Machine-learning-based pricing.

5.1. Challenges

Currently, the biggest challenge in the data trading market is the lack of a unified pricing benchmark, especially for machine learning, so all kinds of data trading mechanisms are not available in practice. To solve the biggest problem, the pricing benchmark of the dataset is essential, such as MNIST, CIFAR-10, CIFAR-100, and ImageNet et al.

Another challenge is how to design an efficient and appropriate incentive mechanism for decentralized machine learning. Thanks to the incentive mechanism on distributed machine learning are still in its early stages and despite the apparent opportunities it offers both federated learning and other machine learning, there exist several critical challenges in data pricing. First of all, data pricing methods are hard to evaluate empirically because it is difficult to distinguish the error of the model from the error of the data method explaining the machine learning model. Besides, for this reason, the state-of-the-art data pricing methods are often qualitative, based on the contributions of the produced machine learning model. To develop better quantitative data pricing methods for the evaluation of data value, we will need to define the goal that an ideal data pricing method should achieve, in terms of different methods that might be suitable for different machine learning tasks.

5.2. Futures Research Directions

Currently, due to slightly different data pricing formulations, lack of definition of the utility function for the variety of existing Machine learning models and no common benchmark,

comprehensive data pricing is not available. Besides, the data pricing issues of the key task on distributed machine learning over time have not been completely addressed in the literature. However, in the machine learning scenario, using game theory and mechanism design tools will be an effective way to solve the data pricing problem. So, the research about data pricing for machine learning is still one of the research focuses on the field of artificial intelligence and big data. Moreover, understanding why complex Machine Learning Model makes a precise prediction can be as crucial as the big data value in many applications.

A research solution of data pricing based on federated learning as illustrated in figure 4.

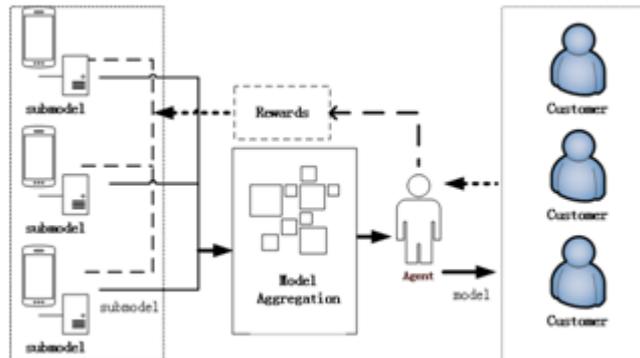


Figure 4. Overview of the data pricing based FL

For centralized machine learning settings, gradient-based attribution methods [1] can be used to explain deep neural networks (DNN) and to evaluate data value, including global attribution methods and local attribution methods. A model aggregating with attention mechanism [7] maybe be used to evaluate the client data value by computing the contribution of distributed client models to the global model during server aggregation. Considering heterogeneous federated learning, use transfer learning and knowledge distillation [14] to develop a universal data pricing framework for distributed machine learning, which enables each participant can gain rewards through owning their private data. Exploring the transferability between heterogeneous [31] datasets sheds light on their intrinsic data value, and consequently enables knowledge transfer from different datasets to the machine learning model to valuation the value of training datasets the latter.

The reliable and efficient operation of distributed machine learning relies on the cooperation of edging nodes, which are the organizations or individuals with different optimization goals. There exist conflicts between individual goals and the overall system objectives. The rational and selfish behaviors of edging nodes would lead the system to an anarchic state and degrade system performance. Therefore, we need to design efficient mechanisms to incentives the cooperation among edging nodes guaranteeing that distributed machine learning runs in an efficient and ordered way. However, designing an appropriate incentive mechanism for machine learning is a complex and challenging task. Intuitively, data is essential in many AI application scenario, and this data may be available with strategic players who need not reveal it truthful to the (AI) system designer unless there are offered proper incentives.

6. CONCLUSIONS

In this survey, we present an overview of the progress of data pricing in data transactions. In particular, we introduce the research progress of the game theory and mechanism design in distributed machine learning in detail, analyze the challenges faced, and look forward to future research directions.

Federated learning, as a type of distributed machine learning, has promising future development, but the challenges it faces are also huge, and these challenges are exactly the direction of a large number of researchers in federated learning.

REFERENCES

- [1] Marco Ancona, Enea Ceolini, Cengiz Oztireli, & Markus Gross, (2018) Towards a better understanding of gradient-based attribution methods for deep neural networks. In 6th International Conference on Learning Representations (ICLR 2018).
- [2] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konecny, Stefano Mazzocchi, H Brendan McMahan, et al., (2019) Towards federated learning at scale: System design. arXiv preprint arXiv:1902.01046.
- [3] Lingjiao Chen, Paraschos Koutris, & Arun Kumar, (2019) Towards model-based pricing for machine learning in a data marketplace. In Proceedings of the 2019 International Conference on Management of Data, pages 1535–1552. ACM.
- [4] Nicholas R Clayton & Hussein Abbass, (2019) Machine teaching in hierarchical genetic reinforcement learning: Curriculum design of reward functions for swarm shepherding. arXiv preprint arXiv:1901.00949.
- [5] Amirata Ghorbani & James Zou, (2019) Data shapley: Equitable valuation of data for machine learning. In International Conference on Machine Learning, pages 2242–2251.
- [6] Zehong Hu, Yitao Liang, Jie Zhang, Zhao Li, & Yang Liu, (2018) Inference aided reinforcement learning for incentive mechanism design in crowdsourcing. In Advances in Neural Information Processing Systems, pages 5507–5517.
- [7] Shaoxiong Ji, Shirui Pan, Guodong Long, Xue Li, Jing Jiang, & Zi Huang, (2019) Learning private neural language modeling with attentive aggregation. In International Joint Conference on Neural Networks (IJCNN).
- [8] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gurel, Bo Li, Ce Zhang, Dawn Song, & Costas Spanos, (2019) Towards efficient data valuation based on the shapley value. arXiv preprint arXiv:1902.10275.
- [9] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nezihe Merve Gurel, Bo Li, Ce Zhang, Costas Spanos, & Dawn Song, (2019) Efficient task-specific data valuation for nearest neighbor algorithms. Proceedings of the VLDB Endowment, 12(11):1610–1623.
- [10] Jiawen Kang, Zehui Xiong, Dusit Niyato, Han Yu, Ying-Chang Liang, & Dong In Kim, (2019) Incentive design for efficient federated learning in mobile networks: A contract theory approach. arXiv preprint arXiv:1905.07479.
- [11] Jakub Konecny, H Brendan McMahan, Felix X Yu, Peter Richtarik, Ananda Theertha Suresh, & Dave Bacon, (2016) Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.
- [12] Paraschos Koutris, Prasang Upadhyaya, Magdalena Balazinska, Bill Howe, & Dan Suciu, (2015) Query-based data pricing. Journal of the ACM (JACM), 62(5):43.
- [13] Chao Li, Daniel Yang Li, Gerome Miklau, & Dan Suciu, (2014) A theory of pricing private data. ACM Transactions on Database Systems (TODS), 39(4):34.
- [14] Daliang Li & Junpu Wang, (2019) FedMD: Heterogenous federated learning via model distillation. arXiv preprint arXiv:1910.03581.

- [15] Fan Liang, Wei Yu, Dou An, Qingyu Yang, Xinwen Fu, & Wei Zhao, (2018) A survey on big data market: Pricing, trading and protection. *IEEE Access*, 6:15132–15154.
- [16] Shengzhong Liu, Zhenzhe Zheng, Fan Wu, Shaojie Tang, & Guihai Chen, (2017) Context-aware data quality estimation in mobile crowdsensing. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE.
- [17] Nguyen Cong Luong, Dinh Thai Hoang, Ping Wang, Dusit Niyato, Dong In Kim, & Zhu Han, (2016) Data collection and wireless communication in internet of things (IIoT) using economic analysis and pricing models: A survey. *IEEE Communications Surveys & Tutorials*, 18(4):2546–2590.
- [18] Alexander Muschalle, Florian Stahl, Alexander Looser, & Gottfried Vossen, (2012) Pricing approaches for data markets. In *international workshop on business intelligence for the real-time enterprise*, pages 129–144. Springer.
- [19] Satoshi Nakamoto et al., (2008) Bitcoin: A peer-to-peer electronic cash system.
- [20] Chaoyue Niu, Zhenzhe Zheng, Fan Wu, Xiaofeng Gao, & Guihai Chen, (2017) Trading data in good faith: Integrating truthfulness and privacy preservation in data markets. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 223–226. IEEE.
- [21] Dusit Niyato, Mohammad Abu Alsheikh, Ping Wang, Dong In Kim, & Zhu Han, (2016) Market model and optimal pricing scheme of big data and internet of things (IIoT). In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE.
- [22] Dusit Niyato, Dinh Thai Hoang, Nguyen Cong Luong, Ping Wang, Dong In Kim, & Zhu Han, (2016) Smart data pricing models for the internet of things: a bundling strategy approach. *IEEE Network*, 30(2):18–25.
- [23] Praveen Paruchuri & Sujit Gujar, (2018) Fusion of game theory and big data for ai AI applications. In *International Conference on Big Data Analytics*, pages 55–69. Springer.
- [24] Ramesh Raskar, Praneeth Vepakomma, Tristan Swedish, & Aalekh Sharan, (2019) Data markets to support AI for all: Pricing, valuation and governance. *arXiv preprint arXiv:1905.06462*.
- [25] Rainer Schlosser & Martin Boissier, (2018) Dynamic pricing under competition on online marketplaces: A data-driven approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 705–714. ACM.
- [26] Nihar Bhadrish Shah & Dengyong Zhou, (2015) Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. In *Advances in neural information processing systems*, pages 1–9.
- [27] Lloyd S Shapley, (1953) A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.
- [28] Weiran Shen, Binghui Peng, Hanpeng Liu, Michael Zhang, Ruohan Qian, Yan Hong, Zhi Guo, Zongyao Ding, Pengjun Lu, & Pingzhong Tang, (2017). Reinforcement mechanism design, with applications to dynamic pricing in sponsored search auctions. *arXiv preprint arXiv:1711.10279*.
- [29] Yuncheng Shen, Bing Guo, Yan Shen, Xuliang Duan, Xiangqian Dong, & Hong Zhang, (2016) A pricing model for big personal data. *Tsinghua Science and Technology*, 21(5):482–490.
- [30] Adish Singla & Andreas Krause, (2013) Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1167–1178. ACM.
- [31] Jie Song, Yixin Chen, Xinchao Wang, Chengchao Shen, & Mingli Song, (2019) Deep model

- transferability from attribution maps. In *Advances in Neural Information Processing Systems*, pages 6179–6189.
- [32] Pingzhong Tang, (2017) Reinforcement mechanism design. In *IJCAI*, vol-ume 17, pages 26–30.
- [33] Surat Teerapittayanon & HT Kung, (2019) DaiMoN: A decentralized artificial intelligence model network. arXiv preprint arXiv:1907.08377.
- [34] Guan Wang, Charlie Xiaoqian Dang, & Ziyi Zhou, (2019) Measure contribution of participants in federated learning. arXiv preprint arX- iv:1909.08525.
- [35] Yingjie Wang, Zhipeng Cai, Guisheng Yin, Yang Gao, Xiangrong Tong, & Guanying Wu, (2016) An incentive mechanism with privacy protection in mobile crowdsourcing systems. *Computer Networks*, 102:157–171.
- [36] Zehui Xiong, Shaohan Feng, Dusit Niyato, Ping Wang, & Zhu Han, (2018) Optimal pricing-based edge computing resource management in mobile blockchain. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE.
- [37] Qiang Yang, Yang Liu, Tianjian Chen, & Yongxin Tong, (2019) Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):12.
- [38] Haifei Yu & Mengxiao Zhang, (2017) Data pricing strategy based on data quality. *Computers & Industrial Engineering*, 112:1–10.
- [39] Zhenzhe Zheng, Yang Gui, Fan Wu, & Guihai Chen, (2014) Star: strategy-proof double auctions for multi-cloud, multi-tenant bandwidth reservation. *IEEE Transactions on Computers*, 64(7):2071–2083.
- [40] Zhenzhe Zheng, Fan Wu, & Guihai Chen, (2014) A strategy-proof combinatorial heterogeneous channel auction framework in noncooperative wireless networks. *IEEE Transactions on Mobile Computing*, 14(6): 1123–1137.
- [41] Zhenzhe Zheng, Fan Wu, Shaojie Tang, & Guihai Chen, (2015) Aegis: an unknown combinatorial auction mechanism framework for heterogeneous spectrum redistribution in noncooperative wireless networks. *IEEE/ACM Transactions on Networking*, 24(3):1919–1932.
- [42] Zhenzhe Zheng, Fan Wu, Xiaofeng Gao, Hongzi Zhu, Shaojie Tang, & Guihai Chen, (2016) A budget feasible incentive mechanism for weighted coverage maximization in mobile crowdsensing. *IEEE Transactions on Mobile Computing*, 16(9):2392–2407.
- [43] Zhenzhe Zheng, Yanqing Peng, Fan Wu, Shaojie Tang, & Guihai Chen, (2017) An online pricing mechanism for mobile crowdsensing data markets. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, page 26. ACM.
- [44] Zhenzhe Zheng, Yanqing Peng, Fan Wu, Shaojie Tang, & Guihai Chen, (2017) Trading data in the crowd: Profit-driven data acquisition for mobile crowdsensing. *IEEE Journal on Selected Areas in Communications*, 35(2):486–501.