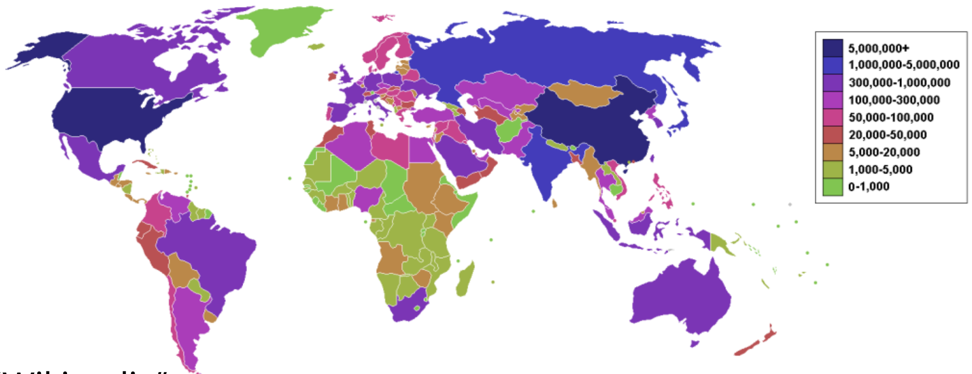
Market-oriented Data Valuation in Smart Grids

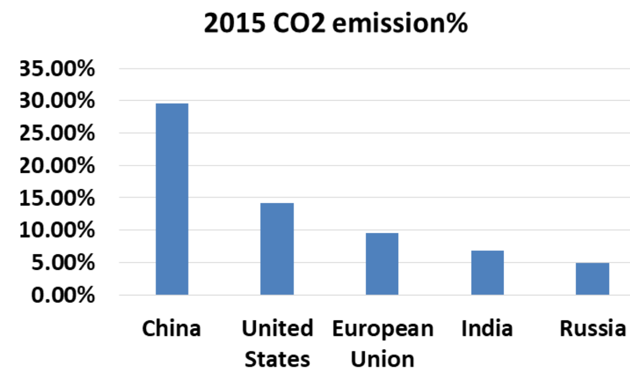
Jianxiao Wang, Ph.D., Assistant Professor
National Engineering Laboratory for Big Data Analysis and Applications

Outline

"3060" Carbon Target of China

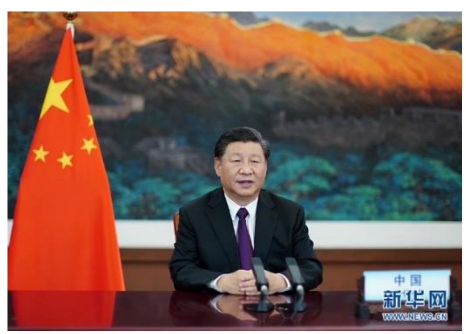
- As the world' s largest coal consumer and coal-derived electricity producer, China has long been suffering massive emissions of GHG and air pollutants.





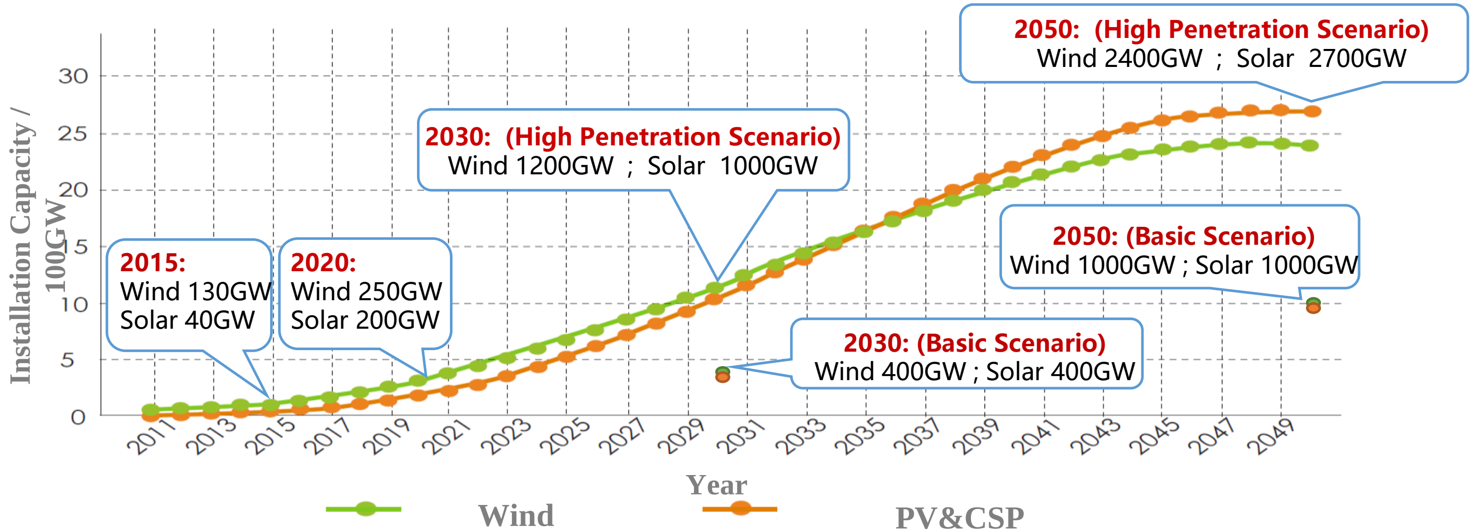
From "Wikipedia"

- On Sep. 22, 2020, China declared "3060" carbon peak and neutrality target, showing an ambitious goal toward carbon neutrality in the future.



High Penetration of Renewable Energy

- It has become a global consensus to develop high penetration and even 100% share of renewable energy toward carbon mitigation.

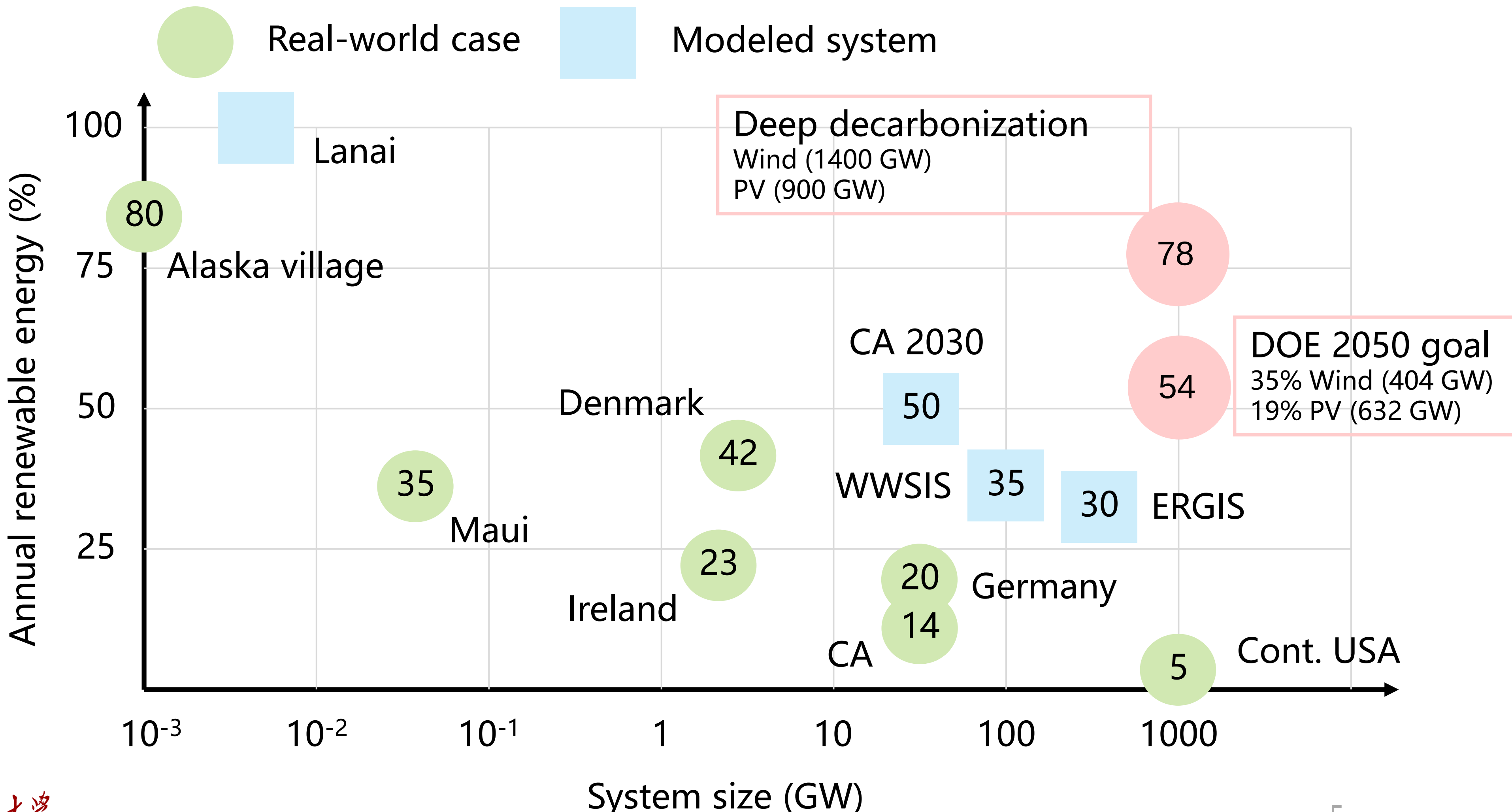


Year		Capacity	Electricity production proportion of renewables without hydro/ %		Electricity production proportion of renewables with hydro / %	
			Whole	Local	Whole	Local
2015		Wind: 130 GW Solar: 40 GW Hydro: 310 GW	3.3	17	22.3	30
2030	Basic Scenario	Wind: 400 GW Solar: 400 GW Hydro: 400 GW	>10	>30	>30	>40
	High Penetration Scenario	Wind: 1200 GW Solar: 1000 GW Hydro: 440 GW	>30	>50	>50	>60
2050	Basic Scenario	Wind: 1000 GW Solar: 1000 GW Hydro: 500 GW	>25	>60	>40	>70
	High Penetration Scenario	Wind: 2400 GW Solar: 2700 GW Hydro: 550 GW	>60	>90	>80	>95

北京大学

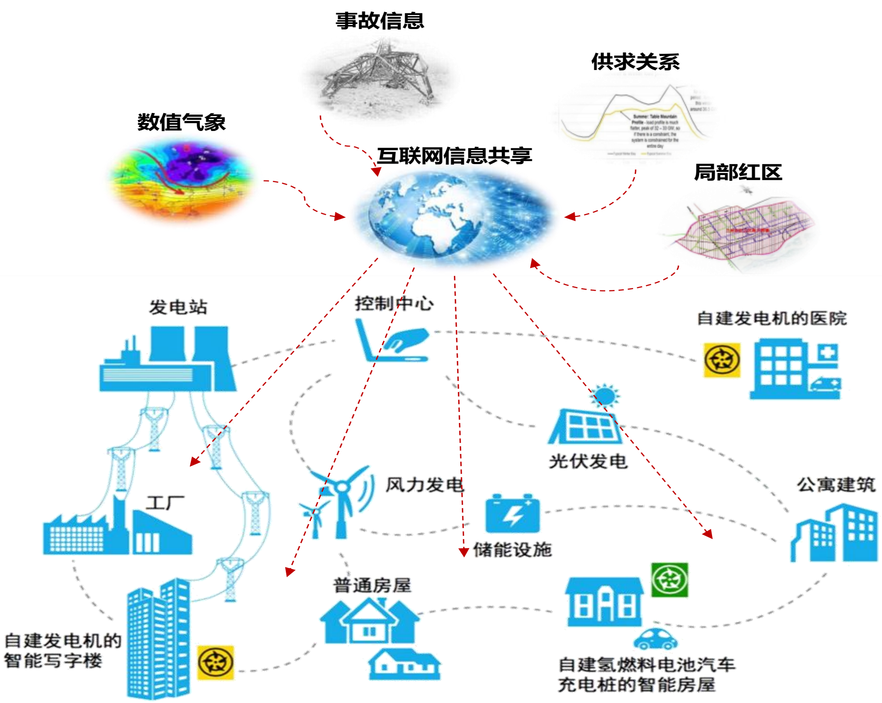
High Penetration of Renewable Energy

- Recent years have witnessed a rapid development of renewable-dominated power systems in many regions across the world.

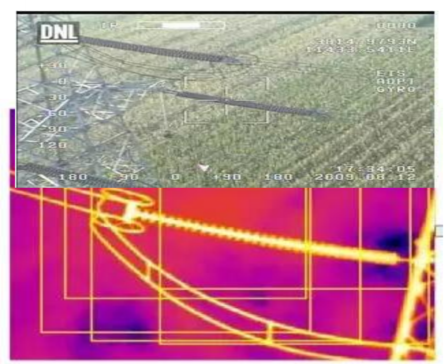


Data Science and Technology

- The importance of data science is increasingly prominent, becoming a supporting technology for energy society.
- Data science and technology enable a more reliable and efficient operation of power grids. There is an urgent need to break through the challenge of mining the value of massive heterogeneous data.



Huawei and China Mobile use image recognition and logistics data to make the world's largest driverless handling system at Tianjin Port work successfully.



The State Grid of China uses image recognition to automate power grid and transmission inspections.

北京大学

Data Sharing and Openness

- China has issued policies that call for accelerating the cultivation of data factor markets:
- promoting the sharing of government data
- enhancing the valuation of social data resources
- strengthening data integration and privacy protection


中华人民共和国中央人民政府
www.gov.cn

简 | 繁 | EN | 注册 | 登录

国务院 总理 新闻 政策 互动 服务 数据 国情 国家政务服务平台

首页 > 政策 > 中央有关文件

收藏 留言

中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见

2020-04-09 19:00 来源：新华社 【字体：大 中 小】 打印

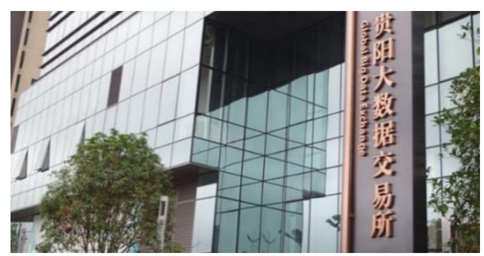
新华社北京4月9日电

中共中央 国务院
关于构建更加完善的要素市场化配置体制机制的意见
（2020年3月30日）

完善要素市场化配置是建设统一开放、竞争有序市场体系的内在要求，是坚持和完善社会主义基本经济制度、加快完善社会主义市场经济体制的重要内容。为深化要素市场化配置改革，促进要素自主有序流动，提高要素配置效率，进一步激发全社会创造力和市场活力，推动经济发展质量变革、效率变革、动力变革，现就构建更加完善的要素市场化配置体制机制提出如下意见。

Beijing International Big Data Exchange

Shanghai Data Exchange

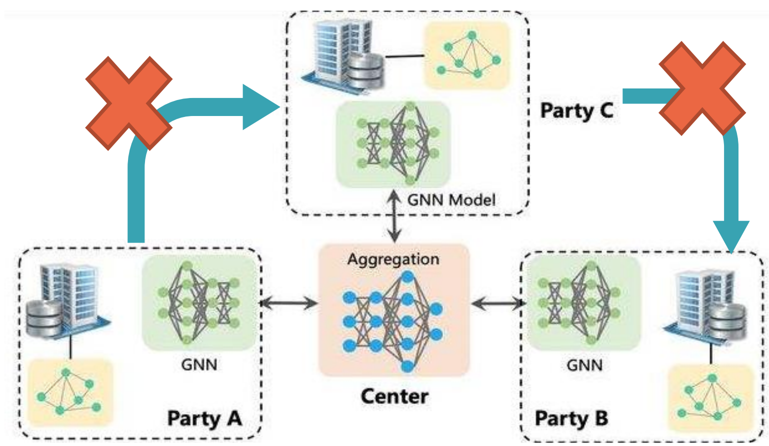


Guiyang Big Data Exchange

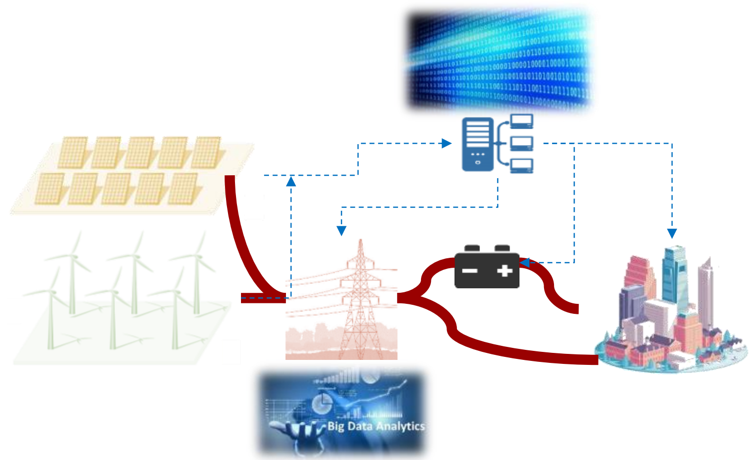

Wuhan East Lake Big Data Exchange

Challenges of Data Sharing

- Data sharing and market trading will bring win-win benefits.
- The lack of an efficient privacy protection technology becomes a major concern for data sharing.
- A standardized data pricing mechanism is absent, which is critical for incentivizing data sharing.



Privacy protection technology

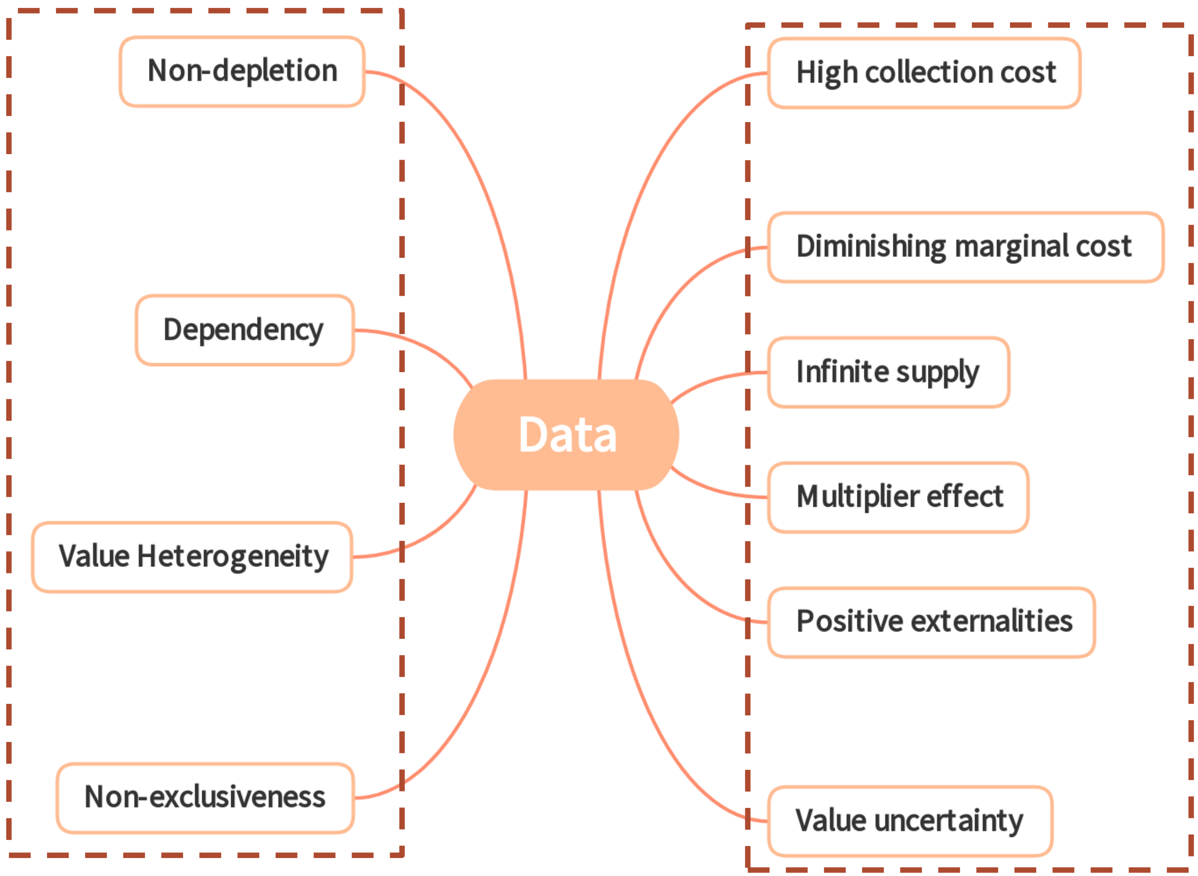


Data pricing mechanism

Outline

Data Attributes

Physical attributes **Commercial attributes**



Data lifecycle

Data Attributes

Security risk

Circulation process

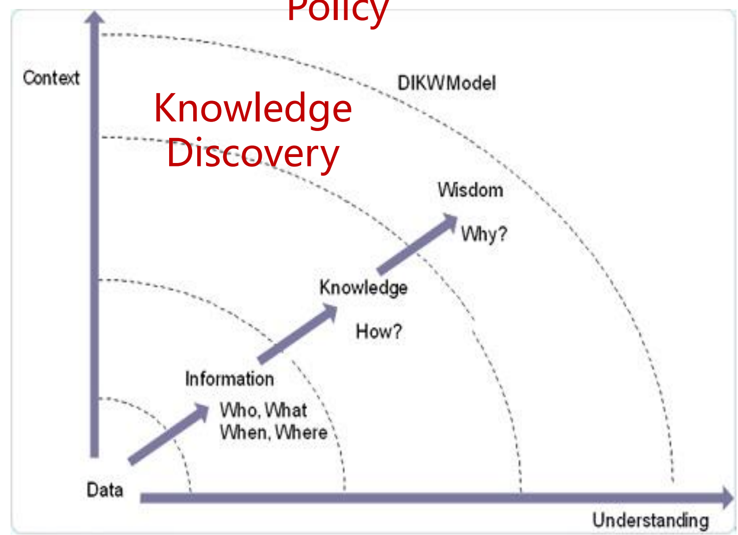
Record → Process → Interpret → Spread

Increasing value

Economic feature

- Non-competitiveness
- Incomplete exclusivity
- Externality

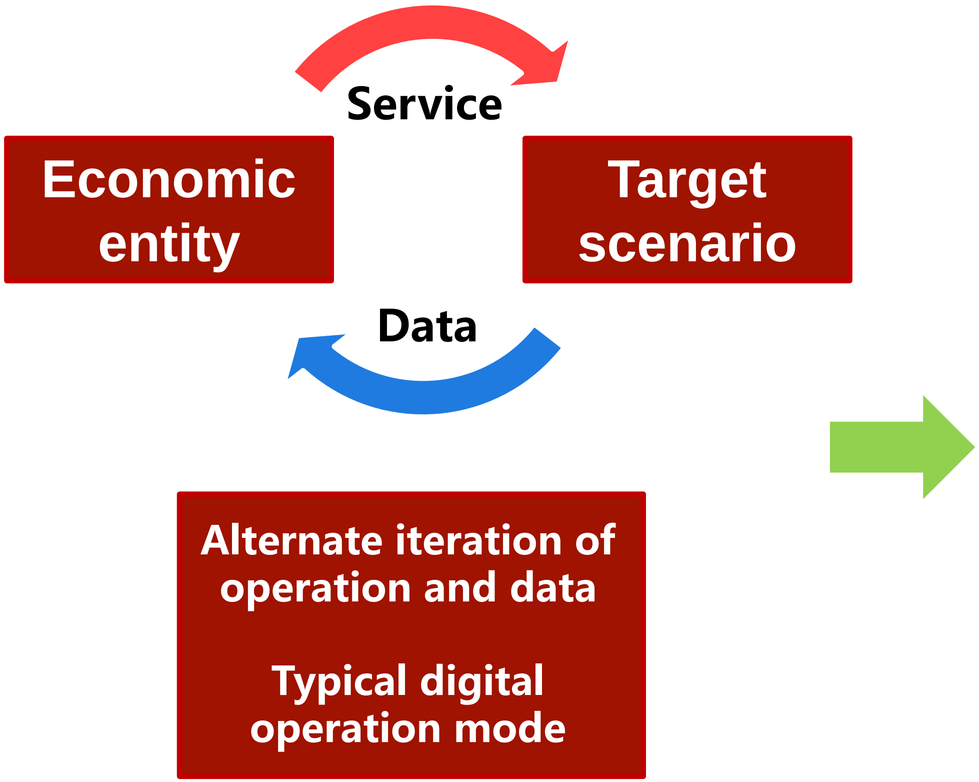
Complex data asset pricing



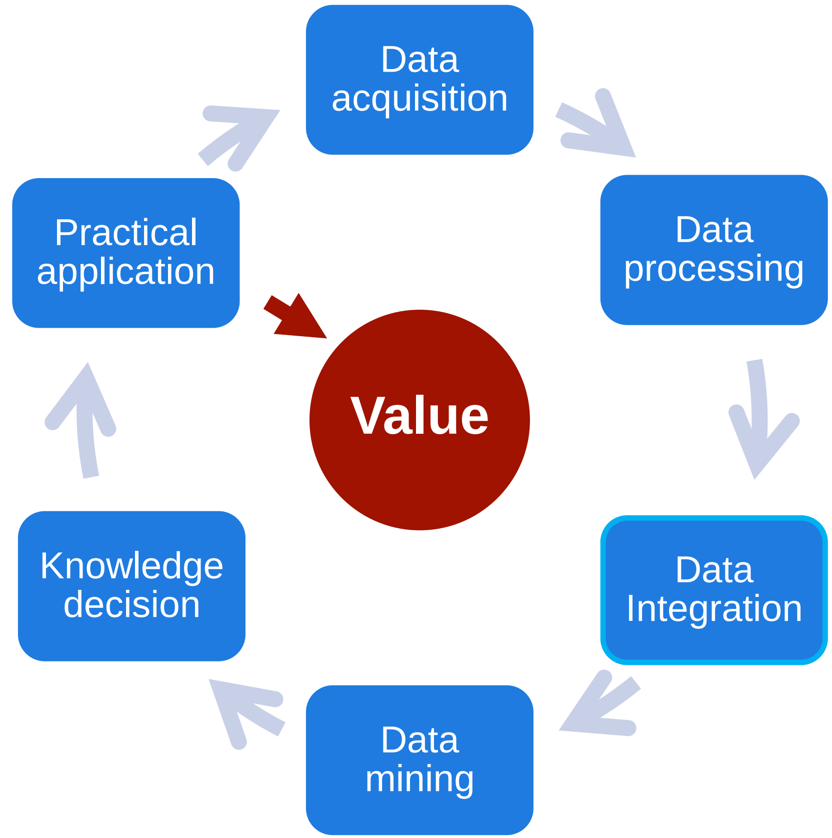
DIKW model

——Russel Ackoff,1989

Data Value

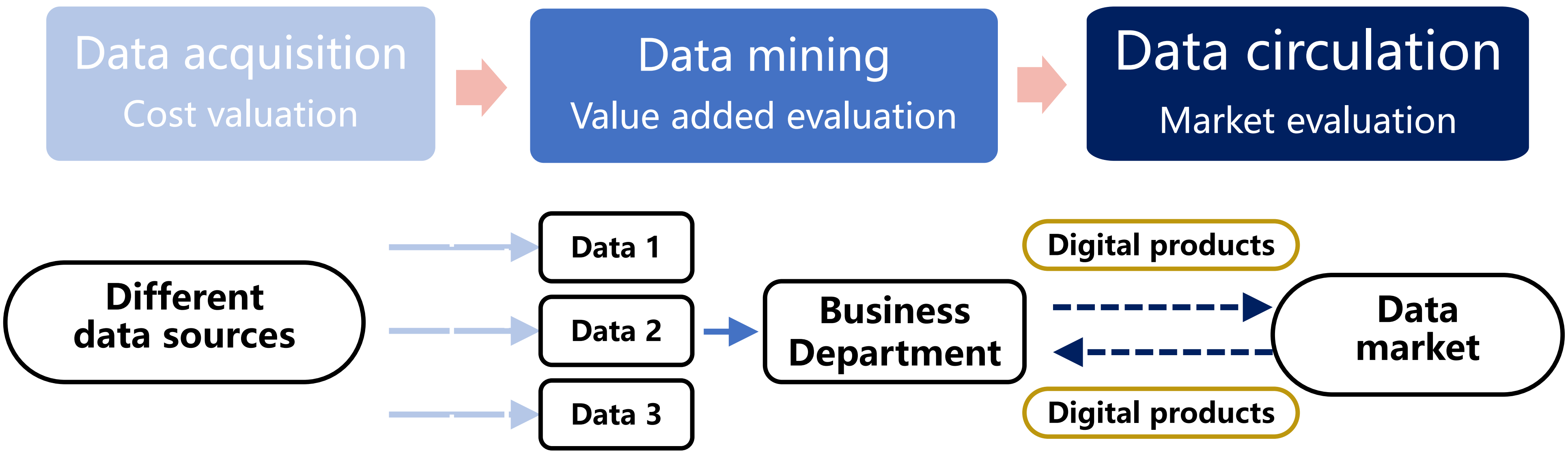


Service
Economic entity
Target scenario
Data
Alternate iteration of operation and data
Typical digital operation mode



Data acquisition
Data processing
Data Integration
Data mining
Knowledge decision
Practical application
Value

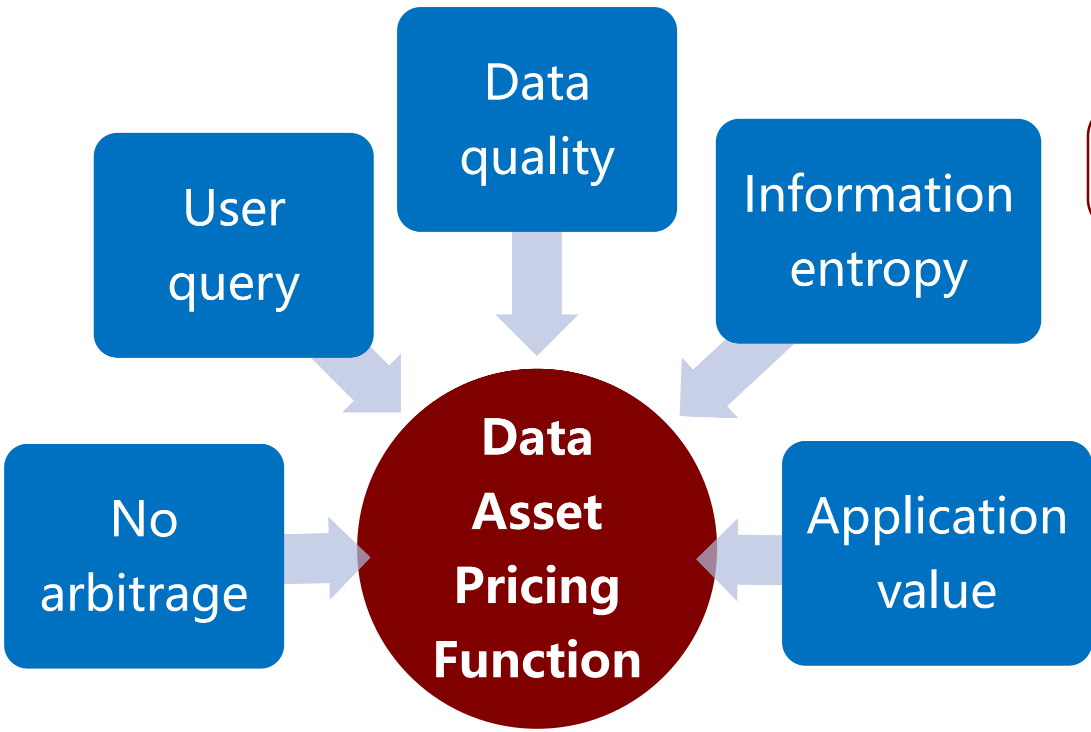
Data Valuation Method



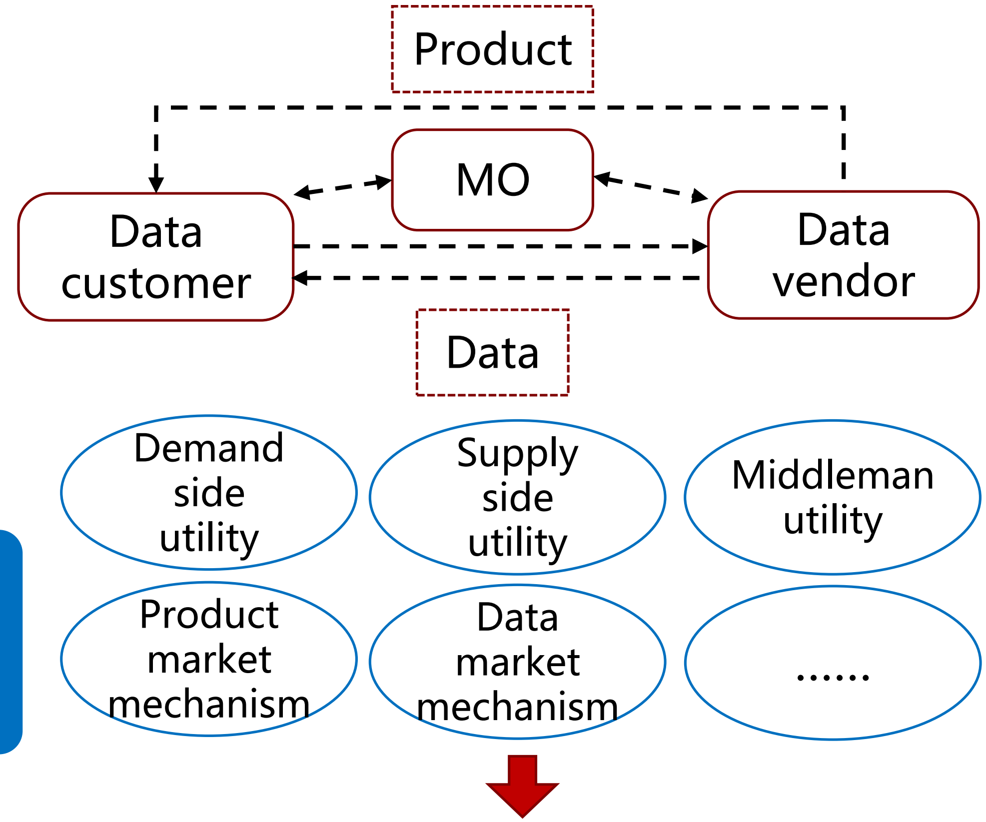
- **Different valuation methods at different stages for different scenarios**
- **Data valuation methods: Cost accounting, utility estimation, market equilibrium**

Examples

Data Asset Pricing Function



Market equilibrium pricing method

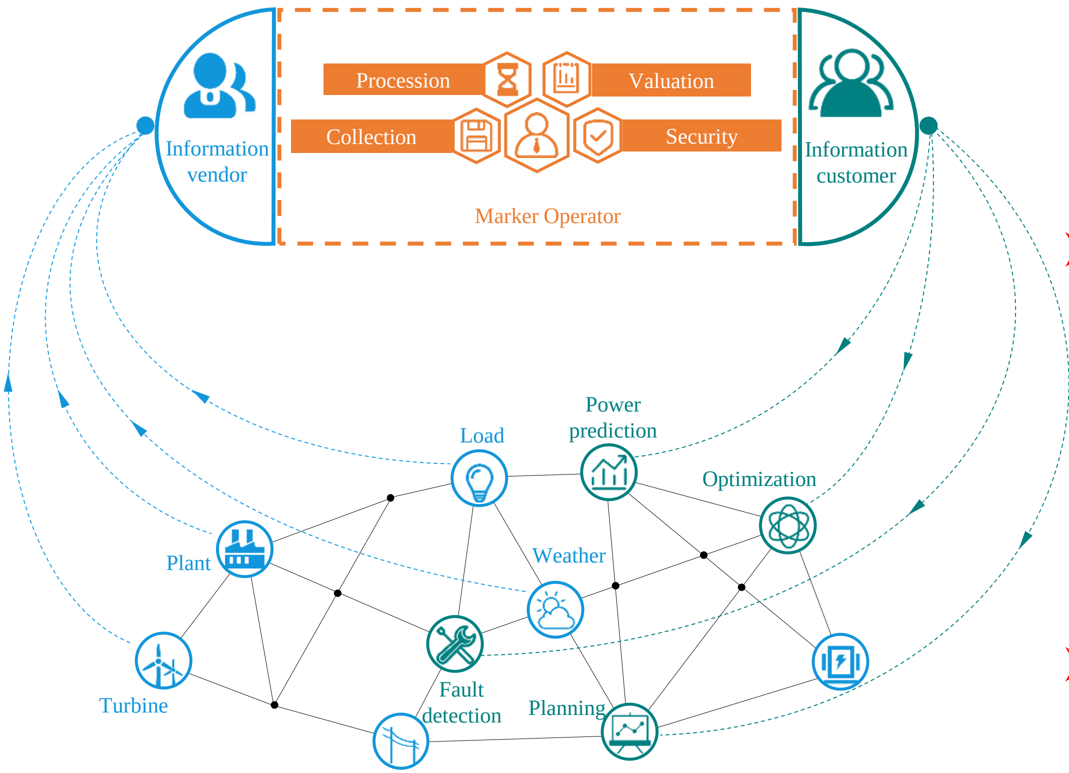


Optimization algorithm
dynamic pricing

Outline

Market Framework

A data market includes three types of entities: data suppliers/vendors, market operator and data customers.



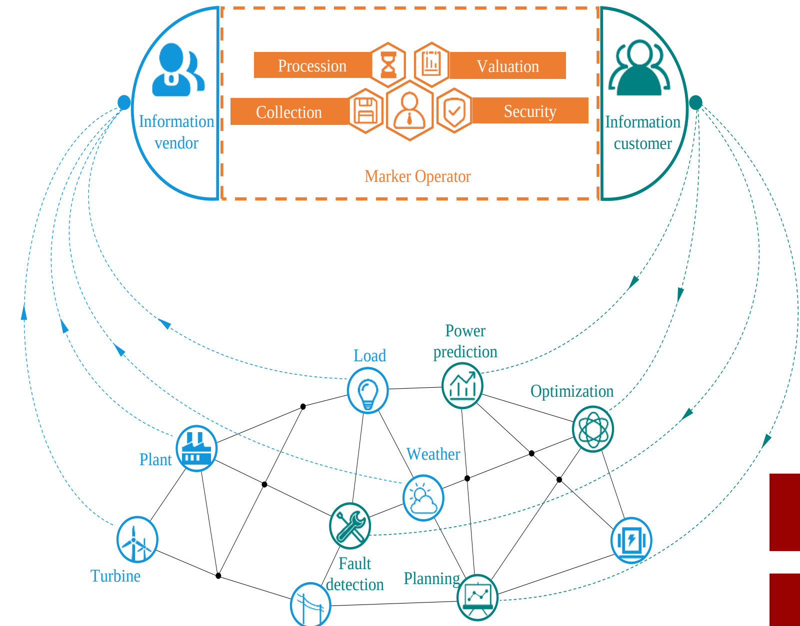
- The data supplier has data collection equipment and can collect a large amount of raw data.

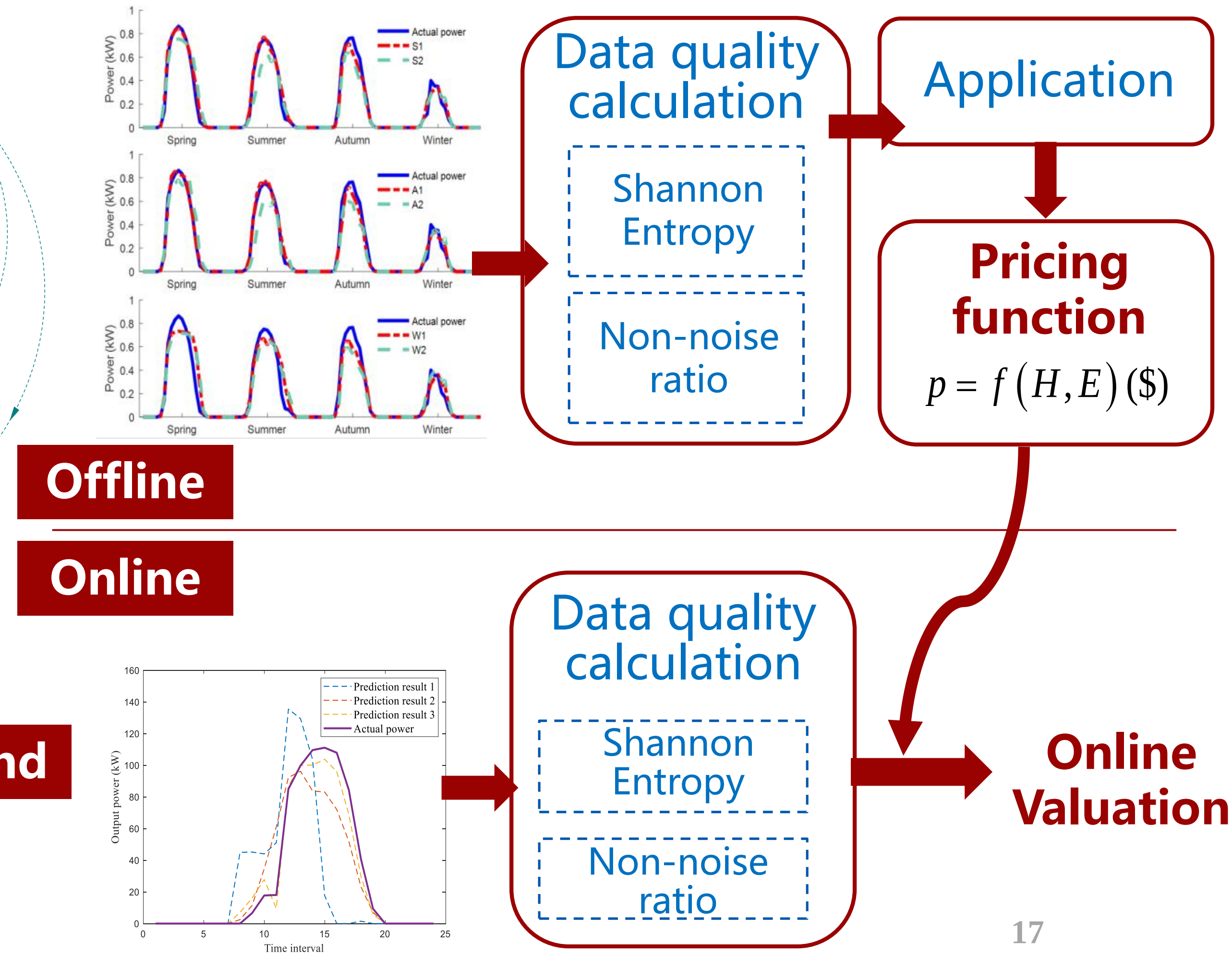
- The market operator is responsible for operating the data market, processing the original data, transforming data into commodities and publishing data prices.

- Data customers have demands and expect to obtain data and create value realization.

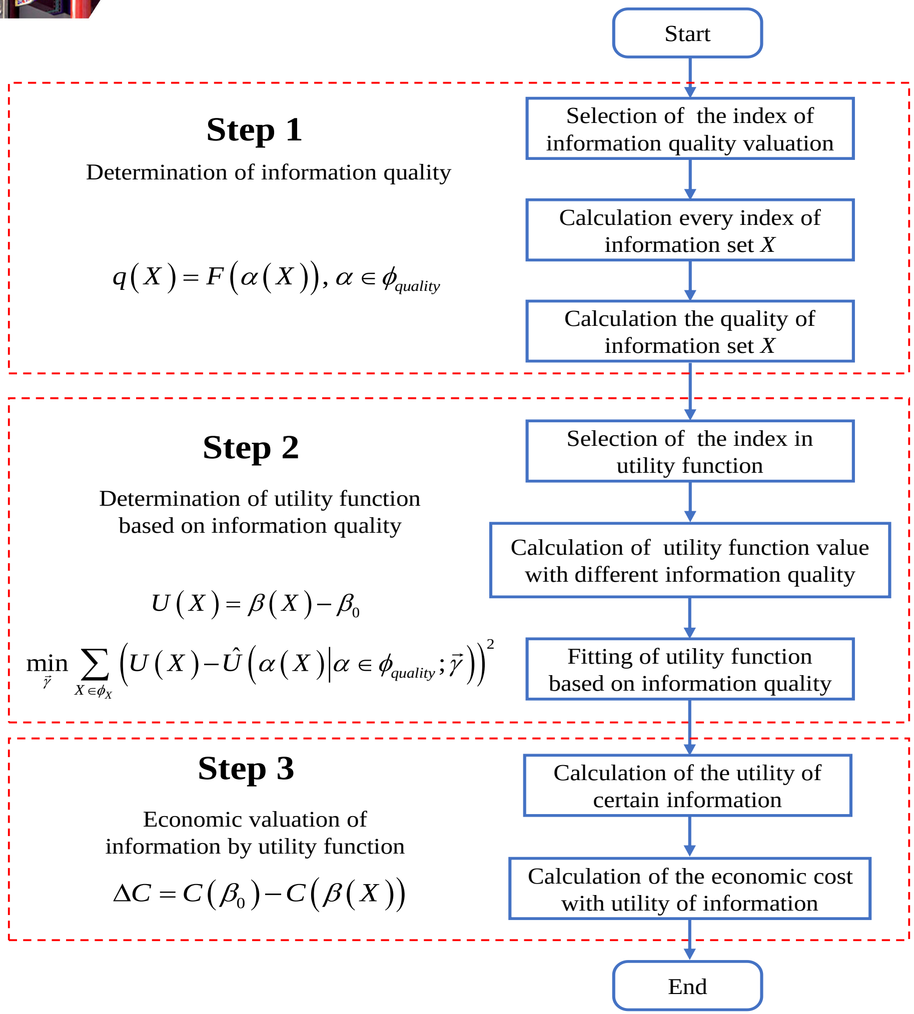
Data Pricing

- **How to quantify the economic value of a package of data assets—Data Pricing Function**





Data Pricing



Generalized data pricing process

(1) Calculate data quality.

Select appropriate indices to calculate information quality, such as data size, non-noise ratio, information entropy, etc.

(2) Calculate the utility of the data.

Quantify the utility of information based on its impact upon a real-world scenario.

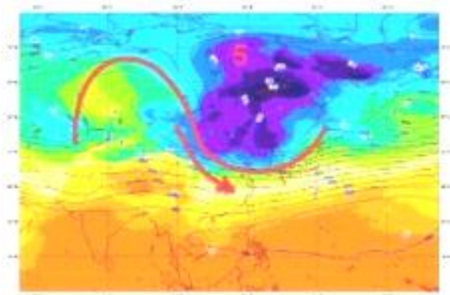
(3) Calculate the economic value of the data.

Select specific applications to quantify the economic value to a physical system. This value can be defined as the reduced cost of physical system operation after obtaining corresponding information.

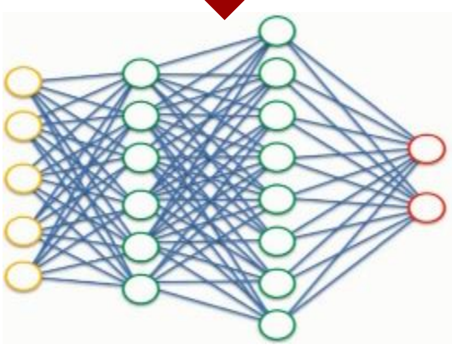
Scenario Description

Based on photovoltaic output prediction, historical meteorological and PV power datasets are priced.

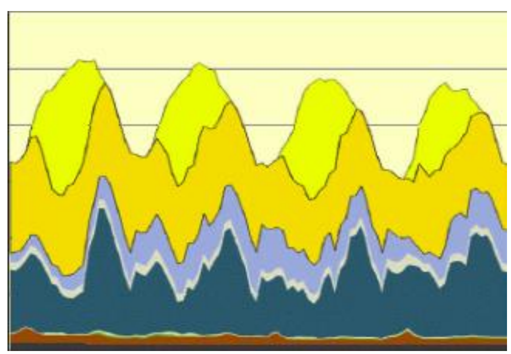
PV data



Meteorological data



Machine learning



Future PV power

Information valuation process

1: Initialization: j=0, N=N_0 (N is the number of information samples)
2: **repeat**
3: Calculation of Shannon entropy: $h_j=h(X_j)$
4: Calculation of non-noise ratio: $\tau_j=\tau(X_j)$
5: Calculation of information quality: $q_j=q(h_j, \tau_j)$
6: $j=j+1$
7: **until** $j>N$
8: Selection of utility function index: Prediction accuracy δ
9: Initialization of utility function: $U(q)$
10: **for** j=0 to N **do**
11: Calculation of utility function: $U_j=\mathrm{U}(q_j)$
12: **end for**
13: Fitting utility function based on U_j
14: Calculation of economic cost $C(\delta)$ by unit commitment model

北京大學

Data Quality

In the renewable energy prediction scenario, Shannon entropy and non-noise ratio are selected for data quality calculation.

Parzen window estimation

$$\phi_{x_{ij}}(x) = \frac{1}{\sqrt{2\pi}\eta} \exp\left\{ -\frac{\left(x - x_{ij}\right)^2}{2\eta^2} \right\}.$$

$$\hat{p}(x) = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \phi_{x_{ij}}(x),$$

Probability calculation

$$\Pr\left[x \in (x_k, x_k + \Delta x)\right] = \hat{p}(x = x_k) \cdot \Delta x,$$

Shannon entropy

$$h(X) = -\sum_{x_{ij} \in X} p\left(x_{ij}\right) \log_b p\left(x_{ij}\right),$$

Non-noise ratio

$$\tau(X) = \frac{N(X) - Z(X)}{N(X)},$$

Quality calculation

$$q(X) = h(X) \cdot \tau(X),$$

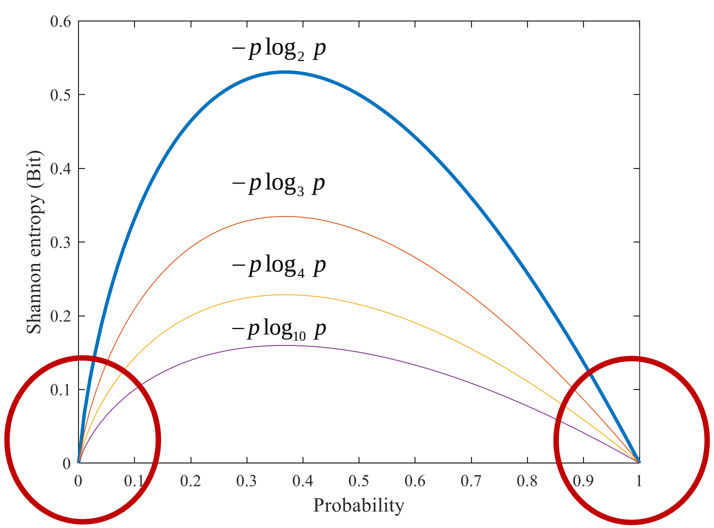
Data Quality

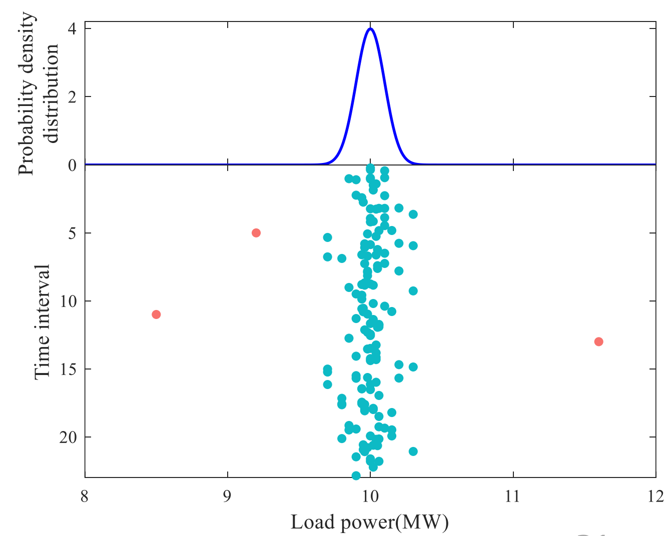
Practical interpretation of Shannon entropy

Suppose that there is a piece of information describing an event A, and the probability of event A is "p(A)" . Shannon entropy is used to evaluate the amount of information available, and judge useless information.

Useless information 1: The information describing an event which always occurs, whose probability is close to 1. This information is common sense.

Useless information 2: The information describing a rare event which hardly happens. This information is far from reality.





Utility and Economic Value Calculation

PV power prediction

Neural network/SVM models are selected.

Input : historical PV power, meteorological data.

(1) Training Stage: feature extraction of data and training of prediction model.

(2) Prediction stage: the trained model is used for prediction, and finally the prediction results are generated.

Accuracy (data utility):

$$\delta = 1 - RMSPE = 1 - \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{P^{PV}-P^{FPV}}{P^{PV}}\right)^{2}},$$

Unit commitment

Objective Function:

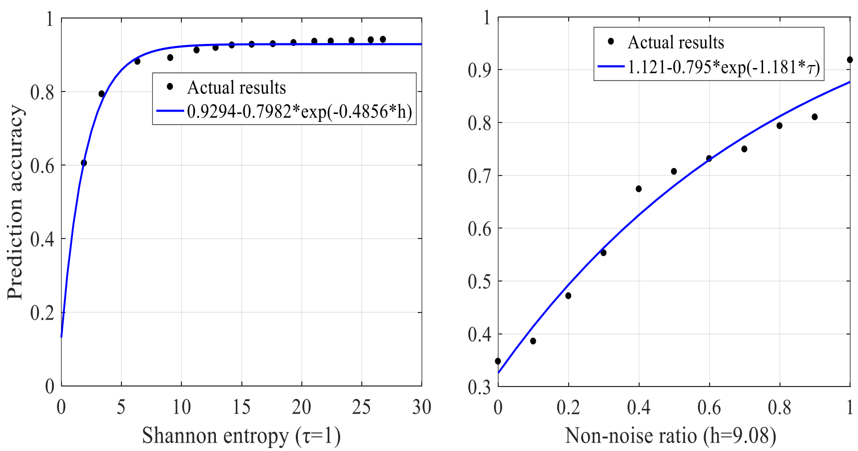
$$\min_{\mathbf{P}} \sum_{t=1}^{T}\sum_{i=1}^{N_G}\left[\alpha_{i,t}C_i^U+\beta_{i,t}C_i^D+\sum_{s=1}^{N_S}\pi_s F_{c,i}\left(P_{i,t,s}^{G}\right)\right] + \sum_{t=1}^{T}\sum_{i=1}^{N_{PV}}\sum_{s=1}^{N_S}\pi_s C_{PV}^{\text{Penalty}}\left(P_{i,t,s}^{PV+}+P_{i,t,s}^{PV-}\right),$$

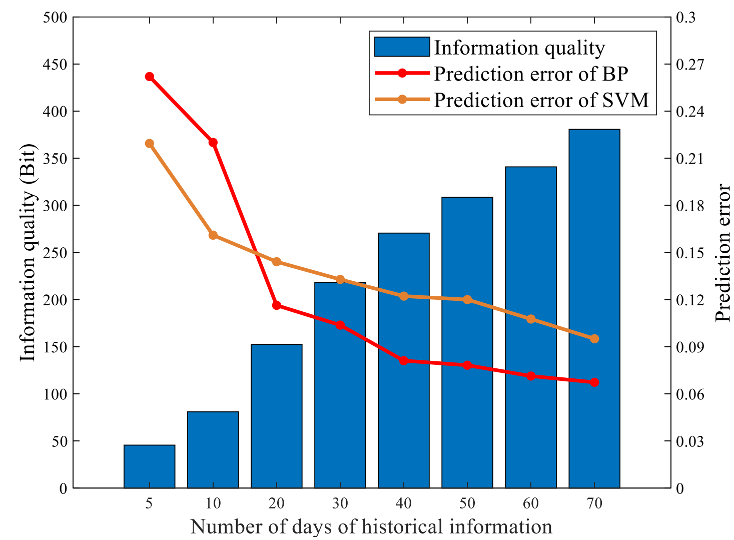
Constraints:

- Output limit of generators
- Starting-up & shutting-down limit
- Ramp rate limits of generators
- PV output limit
- Power balance constraint
- Line capacity limit

Case Studies

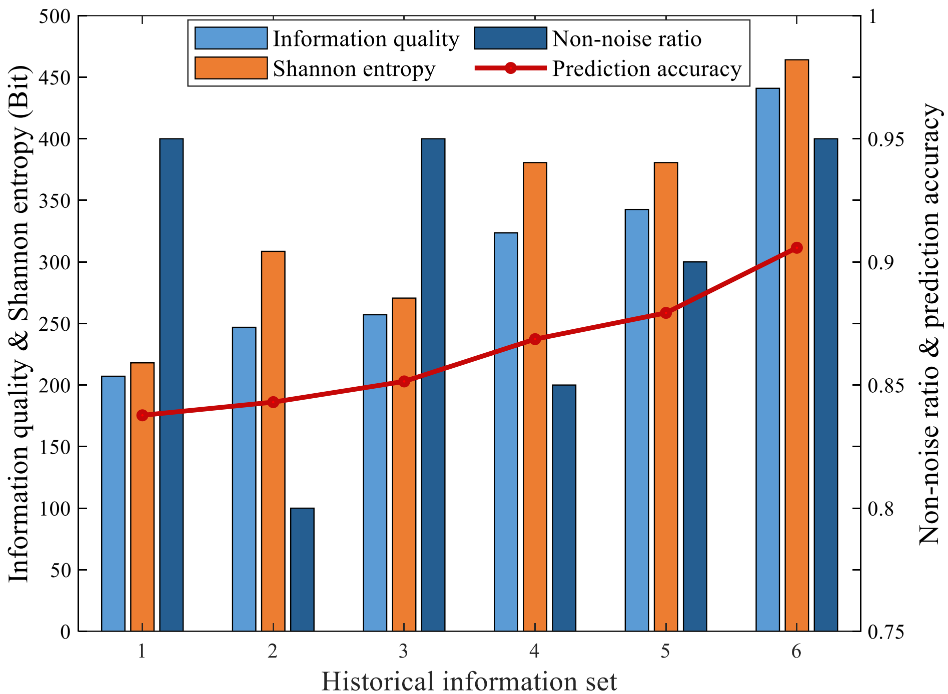
- The prediction accuracy (utility) is regressed in an exponential form with respect to the Shannon entropy and non-noise ratio (data quality) of different datasets.
- In contrast to SVM, a larger dataset helps neural network to achieve a lower prediction error.

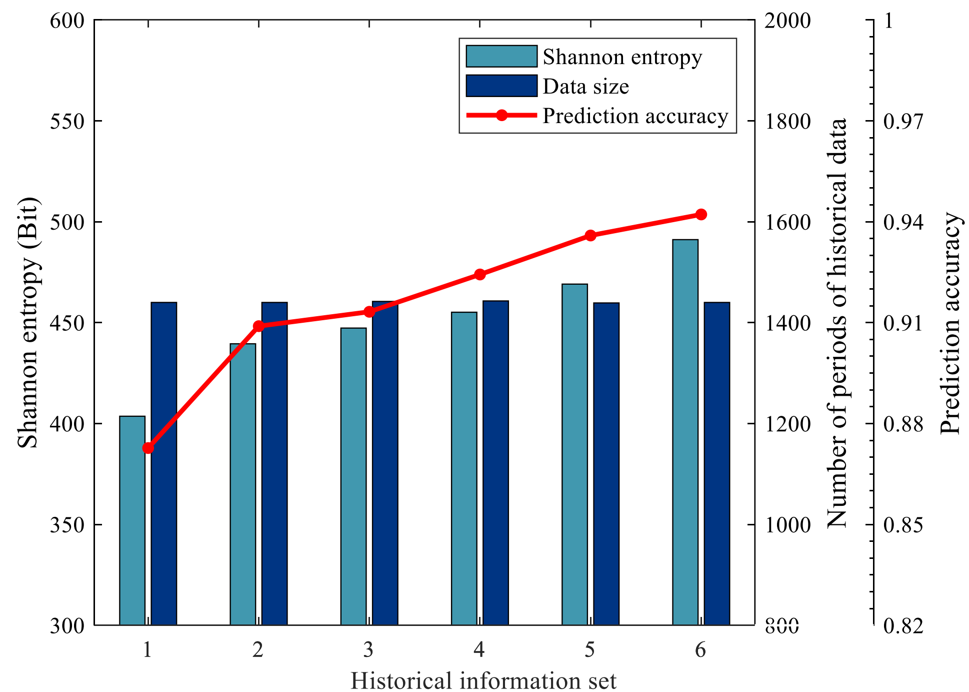




Case Studies

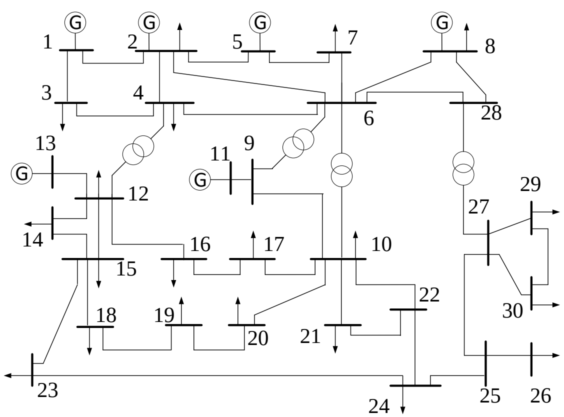
- For online tests, prediction accuracy shows a strong positive relation with information quality measured by both Shannon entropy and non-noise ratio.
- In contrast to the existing studies measured by data size, Shannon entropy demonstrates a better performance in capturing the relationship with accuracy.



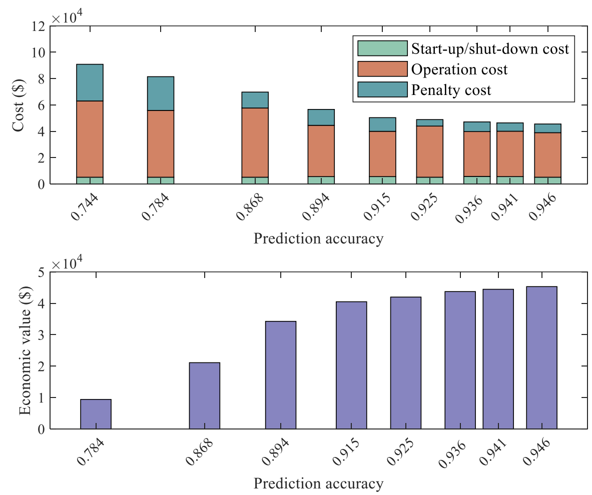


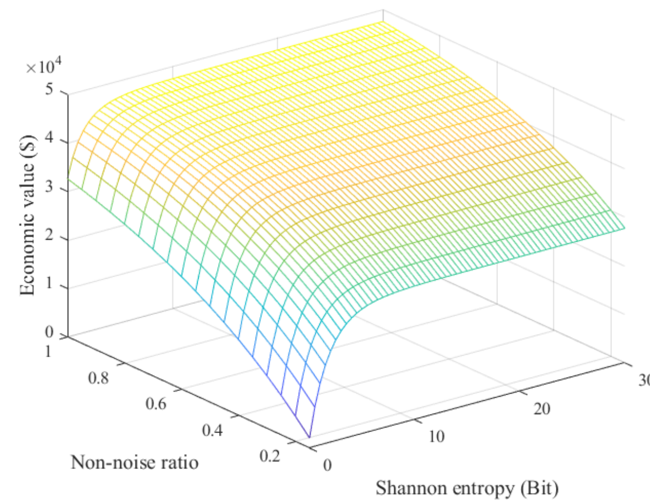
Case Studies

- The IEEE-30 bus system is selected to calculate data economic value.
- The forecast results are input into the unit commitment model, generating the day-ahead scheduling while obtaining the system operation costs.



IEEE 30-bus system





- The system cost decreases with the increase of photovoltaic prediction accuracy. The cost reduction can be defined as the economic value of data.
- However, the marginal data value declines with the increase of accuracy.

Outline

Discussions

- For the same dataset, we are curious about the economic value of different algorithms.
- In real-world cases, the computation burden of different algorithms can lead to the diversity in load consumption of servers.
- Utility efficiency=data economic value/computation time ($/s)

Data Center Server Model

$$E_{host} = P_{host} T^{cal}$$

$$P_{host}(R) = P_{host}^{idle} + R \cdot (P_{host}^{\max} - P_{host}^{idle})$$

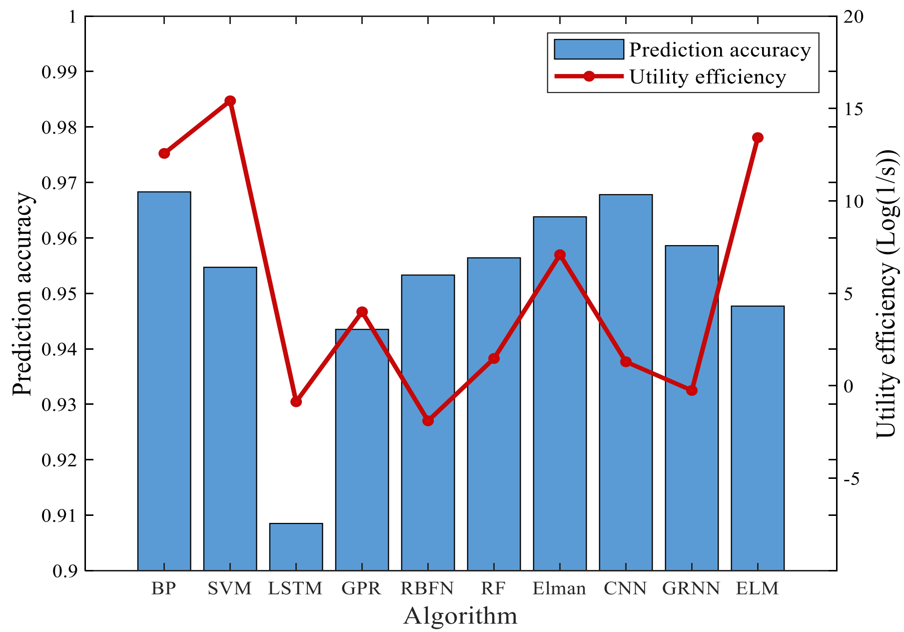
$$E_a = P_a T_a^{cal}$$

$$P_a = R_a \cdot (P_{host}^{\max} - P_{host}^{idle})$$

Quantifying the resources occupied by the algorithm

$$T_a^{Cal} = t_a^{train} + t_a^{predict}, a \in N_A$$

$$R_a = M_a / M_{total}$$



北京大學
PEKING UNIVERSITY

Thanks！

Jianxiao Wang, Ph.D., Assistant Professor
Email: wang-jx@pku.edu.cn