

What Is Big Data

Eizo Kinoshita and Takafumi Mizuno

Abstract This chapter consists of three parts. In first section, we describe what Big Data are. We explain concepts of Big Data, and how it arose. Big Data affect scientific schemes. We mention limitations of predictions by use of Big Data, and a relation between Big Data and hypotheses. And we note that electric power of Big Data systems. In next section, we describe necessity of Big Data. This is a view from aspects of macroeconomics. In service science capitalism, measurements of values of products need Big Data. Service products are classified into stock, flow, and rate-of-flow-change. Immediacy of Big Data implements and makes sense of each classification. And we provide a macroeconomic model with behavioral principles of economic agents. The principles have mathematical representation with high affinity of correlation deduced from Big Data. In last section, we provide an explanation of macroeconomic phenomena in Japan since 1980 as an example of use of the model.

Keywords Big Data · Macroeconomics · Thetical and Antithetical economics

1 What Is Big Data

People have found a vague uneasiness in every field. This is that we will not be able to treat floods of data made by information communication technologies. Big Data is a concept of data which is “impossible to treat by use of traditional technologies”, and is “provided by information communication technologies.” There are some definitions of Big Data, we adapt 3 V definition in this chapter; Big Data are data which are

E. Kinoshita (✉) · T. Mizuno
Meijo University, Kani, Gifu, Japan
e-mail: kinoshit@meijo-u.ac.jp

T. Mizuno
e-mail: tmizuno@meijo-u.ac.jp

1. large in quantity (Volume),
2. traded high speed (Velocity),
3. formed in many style (Variety).

Big Data is impossible data for traditional technologies, but it would be resources if we can overcome impossibility. A term “data” has two different meanings

- information which are obtained experiment or survey for problem solving or decision making,
- information stored in computers.

Big Data have these two meanings simultaneously.

Computer scientists have predicted arising of Big Data earlier than 2001. In those days, broadband wires were spreading ordinary homes, number of portable phones was increasing, and B2C commerce had just realized. Technological innovations miniaturized computer terminals, reduced costs of communications, accelerated communication speed, and released computer networks to open societies. Computer networks became our common property. People can use Web, search information, and send their message to the world. Corporations find values in searching information. Google added a value which is PageRank to the searching information, and Google monopolize portals on the Web. Around 2005, people want environment which they can access to information whenever or wherever they are. This demand is realized by cloud infrastructures. Corporations provide services on the cloud infrastructures at free price. And now, people use Web searching, mail, social networks, and e-commerce. It has become quite common. But people perceive a danger that information become bigger as they can treat. Around 2010, a term Big Data arose to societies.

Big Data represent projections of things on real world, thinking of people, results of calculations of computer. To say concretely, they are numerical values, texts, images, movies, sounds, programs, and so on. They are coded by any rules, and stored in storages of computers. Computers can proceed any procedures to the coded things. The procedures are operations on binary digits essentially. So, infrastructures of Big Data are computer systems, and theoretical backgrounds of Big Data is computer sciences and information communication technologies.

Computer is a system consists of four parts: arithmetic units, memories, I/O interfaces, and storages. Data are stored in storages. When a computer processes some procedures, computers load data and programs from storages to its memories. Arithmetic unites process data in the memories as calculation on binary digits. Results of the processing are displayed to operators via I/O interfaces, or stored into storages.

Storage accesses occupy almost all times of processing with large volume data. Access times between arithmetic units and memories are nanoseconds order. While access times between memories and storages are micro seconds or milliseconds order. Or the speed of data transfer between memories and storages is about gigabits per seconds. Now, the speed of data transfer on Web is about gigabits per seconds or hundreds gigabits per seconds. It means that storage access times and network access

times are same order. We do not mention which data are on storages or beyond networks. Network speeds give us permeability of data accesses on networks and realize using Big Data.

Infrastructures of Big Data consist of many computer nodes. When we process Big Data by use of more than ten thousand nodes, we need specific software technologies and hardware technologies. For examples, Google constructs Big Data systems which consist of about million nodes. And Google also provides Big Data infrastructures: Google File System, MapReduce, BigTable.

1.1 Predictions by Use of Big Data

Theories used to be our experiences which are arranged economically. Because costs of memorizing are higher than costs thinking for humans, we generalize our experiences and avoid memorizing all experiences. Searching of information from vast experiences has high costs for humans, too. So, memorizing has high status in every examination for humans. Until about twenty years ago, there are experts who treat data and sources in every corporation, administrations, and universities. But now, we will change drastically the situations.

- Costs of memory close to zero for person.
- We can test a hypothesis with large data.
- We can use all data at statistics operations.
- We can downloads data which used to be offered by experts.
- We can ask any questions, and we can obtain their answers immediately.
- We can memorize our ideas anytime and anywhere, and we can edit the ideas anytime and anywhere.

A position of the memory in our lives falls extraordinary. Thinking used to supplement to memory, but purposes of the thinking will change. A role of thinking will be that Big Data systems cannot do.

Big Data are large volume data and exist with systems which treat the large volume data. Generally, increasing data volume improves the precision of predictions in scientific field and social sciences. When we introduce Big Data to people, we often tell successful stories of Big Data, too. A prediction flu from social messages on Web by Google, a recommendation system of Amazon, a retail link system of Walmart, and so on. They are successful case studies of improvement of predictions by use of Big Data. There are some factors which lead the cases their successes.

- Attributes or variables which are used in the predictions are concentrated by experts in advance.
- Users do not control their environments by the predictions.
- Environments do not drastically change.

In other words, the cases are predicted well by correlations.

Increasing variables increases the number of data which are required for precise prediction. The increase is proportion to exponential scale of the number of the variables. Although Big Data has enormous data, the volume of data are less than required data by the increasing variables. It is often referred to as curse of dimension.

Generally, predictions by use of Big Data are based on correlations. It means that the predictions does not say causality. The prediction does not tell us input-output relations; variables of the predictions are not causes, and output of the predictions are not results. So, we cannot control outputs of a system by adjusting inputs of the system with the predictions.

Big Data fill environments with enormous data. Predictions by use of Big Data suitable for the environments, and Big Data explain the environments well. But, in areas beyond the environments, the predictions are not suitable. We must common premise between the environments and the areas when we use the predictions. It is difficult to abstract the premise from correlations. The difficulty occurs when the environments drastically change.

1.2 *Big Data and Hypotheses*

Big Data are information which are obtained from systems. The information corresponds results of experiments or observations in scientific fields. A system consists of minute parts, and each parts have relations mutually. If we analyze each parts, we cannot understand the system because of mutual actions. Big Data are obtained from such systems. Because systems are complex, we must collect large volume data. Urban, human, weather, traffic, software, financial trade, and other economic actions are systems. When we combine parts, the complexity of a system appears as new characteristics of the system.

We cannot understand a system by just viewing the system. So we ignore or simplify complex parts of the system for understanding or study of the system. The simplified representations of the system is referred to as model. The model is the abstracted of the system. The model, of course, is differ from the system. But we can understand the system via the model. The model represented by artificial languages or mathematical symbols is referred to as mathematical model. Mathematical model can represent actions of the system by algebraic operations.

Let us consider a mathematical model with any parameters. We can specialize the mathematical model by fill the parameters with any values. The specialized mathematical model is referred to as hypothesis. In statistical fields, targets of tests are the hypothesis.

$$\text{model} + \text{specific parameters} = \text{hypothesis} \quad (1)$$

We can test hypotheses by concrete procedures. A hypothesis which tolerate against various tests are called theory. We can use Big Data for constructing hypotheses, and

for testing the hypotheses. In other words, Big Data can make hypotheses. And the hypotheses are representations of system that we want to understand.

We can understand a system by combining data and model, not only seeing data. Arising Big Data and Big Data systems reduces costs of data, and reduces values of data itself. We treat Big Data as assets. It means, to be exact, our assets are Big Data with model, and hypotheses made by Big Data.

1.3 *Big Data and Electric Power*

On physical aspects, we can grasp Big Data by electric power. Big Data systems consist of thousands computer nodes or million computer nodes. Required electric power is million times as large as one personal computer. Heat produced by the systems are also large. Big Data system is often in data centers. Computer nodes arranged into the data centers. Each data center has huge buildings to cool down computer nodes. Electric power which one computer nodes consumes is hundred watt order. If there are ten thousands computer nodes, required electric power is megawatt order.

If we see services provided by Big Data systems as merchandises, raw materials of the merchandises are electric power. Big Data systems have unique cost structure; we can raise profit rates by reducing electric power for produce one merchandises. Electric power is most important for Big Data systems, because consuming electric power by the system is very large, and because using electric power effectively rises profits directly.

2 **Why We Need Big Data**

We need appropriate use of Big Data to manage our society. If we solve an urban complex problem, the solution makes new issues. Strategic solutions for the issues need large data and data which are obtained without delay. Urbanization and appearance of Big Data change our approaches for the issues. Before appearance of Big Data, we search data for causal relationships. But after Big Data, because of its huge size, we can acquire sufficient correlation to solve the issues. It means that correlation substitutes for causation (Fig. 1).

Information communication technologies reduce sectionalism of governments and give us solutions for urban problems by collaboration of mutual sections. Impor-

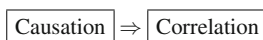


Fig. 1 Relationship in our societies dominated by Big Data

tance of leader who plays role of control tower of each section increases in such mutual society. The leader is CEO or president in corporations, and is prime minister in government. Decision makings design our society. Grand design of society consists of designs of each sector of the society. Each design must be verified that the design is in accordance with the grand design. We need Big Data to construct a strategy by integrating designs and to verify accordance between the strategy and the designs. The society is dominated by economics. Big Data from economic activities are important us for our decision makings or for the construction of the strategy. In this section, we describe necessity of Big Data from aspects of macroeconomics.

In 1980s, researchers of macroeconomics recognized difference between goods products and service products, and they have tried to define what service products are. Now, service products are defined as products that have properties: intangibility, immediacy, variability, perishability, and customer's high satisfaction.

A major premise of macroeconomics is that our world is capitalism. If the world is not capitalism, then every theory of macroeconomics will lost its senses. So, researchers of macroeconomics, managers of companies, or government administrators must consider whether we are in the world with capitalism.

The most important concept of capitalism is fixed price sales. Fixed price sales enable us to run our planned business and guarantee value of capitals.

To enforce fixed price sales without any contradictions on our business, we must measure values of our products precisely. In a word, precise measurements of products provide bases of every index about economics and managements in the world of capitalism; the measurement of values of products is an element forming economics and managements.

For any goods products, we can measure its values relatively easily. Because the goods have physical entities and properties, we can reduce eventually their values to their length, weight, temperature, velocity, or entropy.

On the other hand, we cannot measure values of service products easily. Service products often stand on relations between goods and goods, or between services and services. Relationship is combinations of products, and increasing the number of the combinations makes measurements of values of the products complex. As service products consist of some lower level services, they are developed in high abstraction level far from physical goods products. To overcome the complexity and the distance abstraction level, we need much knowledge of many fields.

In early 2000s, IBM researchers advocated a necessity of "service science" which is a new research filed to construct knowledge systems for service products. We need accumulation of knowledge. It means that we must collect Big Data and extract new theories form Big Data.

We refer to a society in which almost all employees work for service industry as service science capitalism society. In the society, every price value has large amount of information in the background of the value, and the value is detected in high abstraction level far from its physical entity. To fill the gap between abstraction levels, we must learn techniques which reduce from Big Data to a value through experience.

Big Data provide us new measurements for service products, and enable us to classify service products into three services: stock service, flow service, and rate-of-

flow-change service. Stock service is construction of social infrastructures or information infrastructures. Flow service is ordinary everyday service which provided by government administrators and private companies. Rate-of-flow-change service is unusual service.

There is an analogy between physics and economics. In physics, a phenomenon is described in distance, velocity, and acceleration. Establishing the three concepts makes modern physics since 17th century. While economics was made by establishing three concepts: stock, income, and growth rate. In economics, a product is described in the three concepts. Distance, velocity, and acceleration in physics correspond to stock, income, and growth rate in economics, respectively. Distance and stock are measured by some accumulations. Velocity and income are represented in time differentiations. Acceleration and growth rate are represented in twice differentiations. The classification of service products corresponds to the concepts of physics and economics.

The classification presumes that we can trace changes of values of service products every times. It corresponds to time derivative in physics. Immediacy of Big Data provides us feasibility the classification.

When we use Big Data sufficiently, correlation plays important roles in any analyses of economics. So we must build macroeconomic models which we can construct by detecting parameters from correlation deduced from Big Data.

Kinoshita provides a macroeconomic model which is referred to as “Thetical economics and Antithetical economics” [1–4]. That is a rearrangement of theories of macroeconomics into two set; a set of them is Thetical economics and another set is Antithetical economics. If Say’s law is valid in an economic phase in an economic cycle, then the Thetical economics dominates the phase. We feel that we are in normal economy and economic growth in the phase. While if the Keynes’s effective demand is effective in an economic phase, then the Antithetical economics dominates the phase. We feel that we are in depressed economy in the phase. Economic phases dominated by Thetical economy and economic phases dominated by Antithetical economy are illustrated in Fig. 2. Easy to say, Thetical economics represents what prosperity is, while Antithetical economics represents what recession is.

With the macroeconomic model, we can provide behavioral principles of economic agents such as corporations and governments as follows [5, 6]:

- A principle of corporations under Thetical economics
 - Objective function (maximize profits)

$$\max \sum_{j=1}^n c_j x_j \tag{2}$$

- Constraint condition

$$\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, \dots, m \tag{3}$$

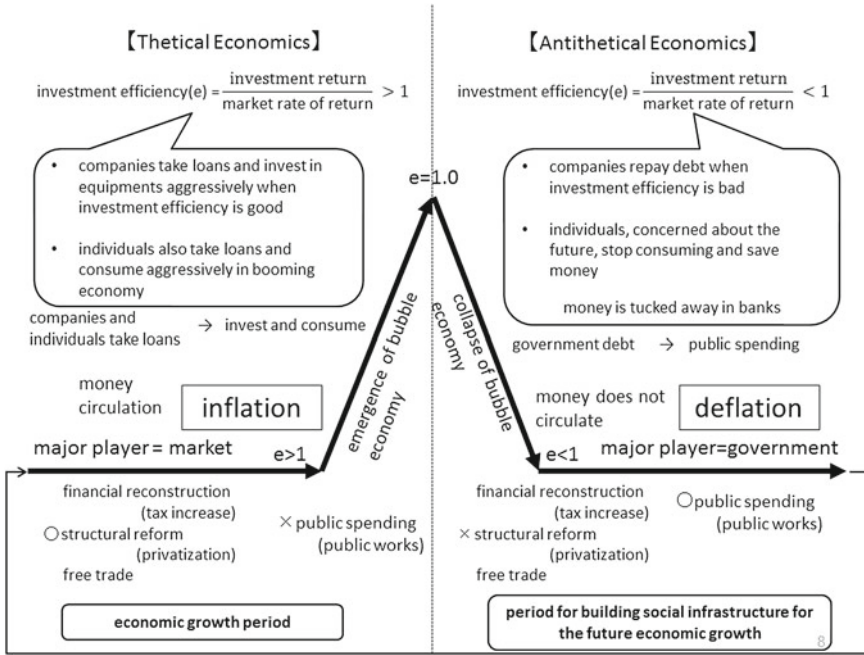


Fig. 2 A economic cycle. The figure is reproduced from a figure which was made by Kinoshita [5, 6]

- A principle of corporations under Antithetical economics
 - Objective function (minimize debts)

$$\min \sum_{i=1}^m u_i b_i \tag{4}$$

- Constraint condition

$$\sum_{i=1}^m u_i a_{ij} \leq c_j, \quad j = 1, \dots, n \tag{5}$$

Following list is correspondence of variables and its meanings.

- x_j The number of units of a product j made by the corporation.
- c_j The amount of profits of one unit of a product j ; $P_j - (1 + r)h_j$, where P_j is price of the product j , r is interest rate, and h_j is cost of the product j .
- a_{ij} Costs in an account subject i to produce the product j for one unit.
- b_i The amount debts of an account subject i .
- u_i Unpaid balance rate for the accounting subject i ; $u_i = 1 - \text{amortization_rate}$.

- A principle of governments under Thetical economics
 - Objective function (fiscal reconstruction)

$$\min \sum_{j=1}^N G_j K_j \quad (6)$$

- Constraint condition

$$\sum_{j=1}^N A_{ij} K_j \leq B_i, \quad i = 1, \dots, M \quad (7)$$

- A principle of governments under Antithetical economics
 - Objective function (fiscal stimulus)

$$\max \sum_{i=1}^M Y_i B_i \quad (8)$$

- Constraint condition

$$\sum_{i=1}^M Y_i A_{ij} \leq c_j, \quad j = 1, \dots, N \quad (9)$$

Following list is correspondence of variables and its meanings.

- K_j A rate of the remainder of national loans for an administrative service j . Increasing the rate increases expenses of the service j .
- G_j Demand for funds as national loans for an administrative service j .
- A_{ij} Satisfaction of a resident i when the government gives the resident one unit of costs of a service j .
- B_i A desiring level of total services of the government for a resident i .
- Y_i The amount of public money to increase satisfaction by one unit for a resident i .

In usual studies of the macroeconomics, economic agents, such as customer, corporations, and governments, are modeled simply. All agents expand their profits, they are well-disciplined, they can acquire all information of markets, and their behavior is rational. The principles, which we provide, give a concrete mathematical model of the rationality.

The behavioral principle is linear equation system. Construction the principle is detecting parameters of the equations. So, the model has high affinity with correlation obtained from Big Data [7].

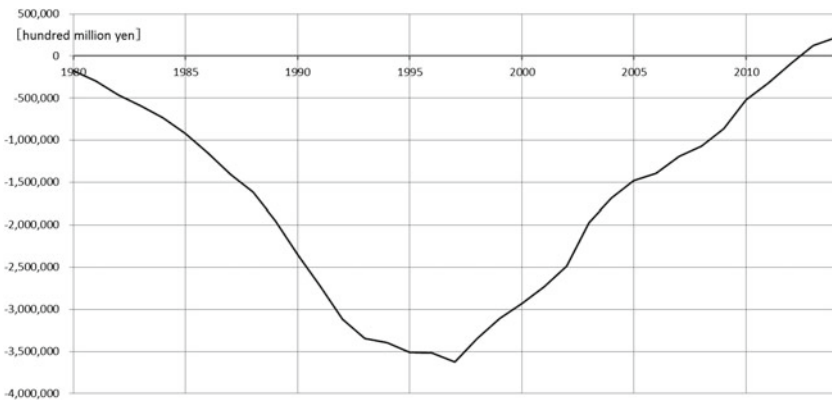


Fig. 3 Financial net worth of non-financial enterprises (total) in Japan from 1980 to 2015. The data is provided by the Bank of Japan

3 The Example of Big Data

As an example, we provide an explanation of macroeconomic phenomena in Japan since 1980 with the model. Let us see Fig. 3, which represents transition of financial net worth of corporations (non-financial enterprises) in Japan. Japan is dominated by Thetical economics before 1995, and is dominated by Antithetical economics after 1995.

Before 1995, corporations increase investments. It is an evidence of behavior of maximization of their profits; the Japanese economy was dominated by Thetical economics. In Japan, Heisei bubble collapse at February 1990. Five years later, Japanese economy was into recession in 1995. Since the year, corporations decrease their debts and increase their savings. It shows a change of behavioral principle of them; the economy is dominated by Antithetical economics.

GDP (Gross Domestic Products) is a macroeconomic index which represents business conditions of the nation. GDP (often denoted in Y) is sum of national consumption (C), national investment (I), governmental fiscal stimulus (G), and trade gap (E).

$$Y = C + I + G + E. \quad (10)$$

Transition of GDP of Japan is shown in Fig. 4. From the change of the index, we can confirm that Japanese corporations do not expand their profits since 1995.

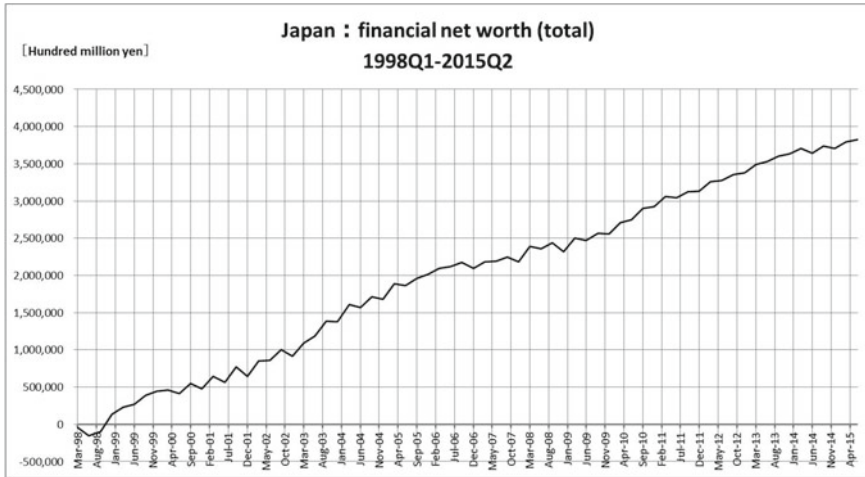


Fig. 4 Nominal GDP of Japan since 1980. The data is provided by the World Bank

4 Conclusions

In this chapter, we describe what Big Data are, and limitations use of Big Data. Experts, of course, use Big Data while considering the properties and limitations of them. We can grasp knowledges deduced from Big Data by paying attention to how a Big Data system treats the properties and overcomes the limitations.

We describe necessity of Big Data with a view from aspects of macroeconomics, and we provide a macroeconomic model with behavioral principles of economic agents. The principles have mathematical representation with high affinity of correlation deduced from Big Data. And we provide an explanation of macroeconomic phenomena in Japan since 1980 as an example of use of the model.

References

1. Kinoshita E (2011) A proposal of primal and dual problems in macro-economics. *China-USA Bus Rev* 10(2):115–124
2. Kinoshita E (2011) Why bubble economy occurs and crashes?-Repeated history of economic growth and collapse. *Chin Bus Rev* 10(2):102–111
3. Kinoshita E (2012) A proposal of thetical economy and antithetical economy-mechanism of occurrence and collapse of bubble economy. *J Bus Econ (Acad Star Publ Co)* 3(2):117–130
4. Kinoshita E, Mizuno T (2015) Trap of economics the world has fallen in a survey of Kinoshita theory in macro-economics. *Eur Sci J*:75–80 (2015). SPECIAL edition
5. Kinoshita E (2015) A proposal of thetical economy and antithetical economy by using operations research techniques. *Eur Sci J* 11(19):29–48 (2015). July 2015 edition
6. Kinoshita E (2015) Thetical and antithetical business management. *J Bus Econ* 6(6):1086–1096 (2015). June 2015
7. Kinoshita E, Mizuno T (2016, in printing) Why we need big data? In: *Proceeding of the 3rd international conference on advances in big data analytics*