BIG DATA

A REVOLUTION

THAT WILL TRANSFORM HOW
WE LIVE, WORK, AND THINK

VIKTOR MAYER-SCHÖNBERGER
KENNETH CUKIER

Big Data

A Revolution that will transform how we live, work and think

Keynotes of Chapter 1-2

Changchun Zhang

Now (Chapter 1)

- Examples:
  - Google took the 50 million most common search terms to identify areas infected by the flu virus.
  - Oren Etzioni predicts if the price of plane ticket is increasing or decreasing in the future, to help customer to determine when to buy the ticket.
- Data become a raw material of business, used to create a new form of economic value.

Letting the data speak

- There is no rigorous definition of big data.
- One way to think about the issue: big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationships between citizens and governments, and more.

- Big data marks the beginning of a major transformation.
- The techniques for collecting and analyzing huge bodies of data will help us make sense of our world in ways we are just starting to appreciate.
- The real revolution is not in the machines that calculate data but in data itself and how we use it.

- Things are speeding up. The amount of stored information grows four times faster than the world economy, while the processing power of computers grows nine times faster.
- What does this increasing mean?
  - By changing the amount, we change the essence. When we increase the scale of the data, we can do new things that weren't possible with smaller amounts.

- Big data is about predictions.
- Big data is not about trying to "teach" a computer to "think" like humans.
- Big data is about applying math to huge quantities of data in order to infer probabilities.
- Big data will change fundamental aspects of life by giving it a quantitative dimension it never had before.

More, messy, good enough

- Big data's ascendancy represents three shifts in the way we analyze information that transform how we understand and organize society:
    - Analyze far more data.
    - Loosen up our desire for exactitude.
    - A move away from the age-old search for causality.

- In this new world we can analyze far more data.
  - Sampling is artificial fetter before the prevalence of high-performance digital technologies.
  - Using all the data lets us see details we never could when we were limited to smaller quantities.

- Loosen up the desire for exactitude:
  - With big data, we'll often be satisfied with a sense of general direction rather than knowing a phenomenon down to the inch.
  - What we lose in accuracy at the micro level we gain insight at the macro level.

- A move away from the age-old search for causality.
  - In a big-data world, we won't have to be fixated on causality
  - Instead, we can discover patterns and correlations in the data that offer us novel and invaluable insights.
  - The correlations may not tell us why happening, but alert us that is happening.

- The big step toward managing data more efficiently came with the advent of digitization.
- Datafication – helps frame the change. It refers to taking information about all things under the sun – including ones we never used to think of as information at all.

- Big data changes the nature of business, markets, and society.

- Value shifted from physical infrastructure to intangibles and now is expanding to data which is becoming a significant corporate asset, a vital economic input, and the foundation of new business models.

- The effect on individuals may be the biggest shock of all.

- Subject-matter specialist have to contend with what the big-data analysis says.

- Big data will force an adjustment to traditional ideas of management, decision making, human resources and education.

- Due to data's vast size, decisions may often be made not by humans but by machines.

- Dark side of big data are considered – Privacy, Individual volition, safeguard the sanctity of the individual.

- New principles are needed for the age of big data.

- Summary:
  - Harnessing vast quantities of data rather than a small portion, and privileging more data of less exactitude, opens the door to new ways of understanding.

  - Big data also overturns the idea of identifying causal mechanism which is self-congratulatory.

More (Chapter 2)

- Using all the data at hand instead of just a portion of it.

- Statisticians have shown that sampling precision improves most dramatically with randomness, not with increased sample size.

- Random sampling has been a hug success and is the backbone of modern measurement at scale.

From some to all

- The concept of sampling no longer makes as much sense when we can harness large amounts of data.

- Sampling loses detail. In many cases, a shift is taking place from collecting some data to gathering as much as possible., and if feasible, getting everything: N = all.

- Random sampling doesn't scale easily to include subcategories, as breaking the results down into smaller and smaller subgroups increases the possibility of erroneous predictions.
- Sampling also requires careful planning and execution.
- Big data is not necessarily big in absolute terms. What classifies them as big data is that instead of using the shortcut of a random sample, data as much of the entire dataset as feasible are used.

- Using all the data instead of a sample isn't always necessary. But in an increasing number of cases using all the data at hand does make sense, and doing so is feasible now where before it was not.

- Sampling will not be the predominant way we analyze large data set. We will aim to go for all the data.