BD 1-1

1

Hello everyone, I am Haiying Che, from Institute of Data Science and knowledge Engineering

School of Computer Science, in Beijing Institute of Technology, in this session, we will discuss Some idea about big data concepts.

2.

Data is the new Oil, this theme was picked up by Clive Humby, a man who has pioneered to use of data with his firm DunnHumby, at the ANA Senior Market Summit at the Kellogg School in 2006. This theme was picked up again by Gerd Leonhard in 2009 and has become part of the mainstream.

Mark Getty is part of the Getty family that sold their oil empire to Texaco. He looked for something to do and decided that there was an opportunity to consolidate the market for photography.

In an Economist article in 2000 (http://www.economist.com/node/288515) he said: “Intellectual property is the oil of the 21st century. Look at the richest men a hundred years ago: they all made their money extracting natural resources or moving them around. All today’s richest men have made their money out of intellectual property.”

3

And we say Data is like Crude oil…., because normally, original data cannot be used directly, they must be processed before really being applied.

Like the crude need to be refined before use.

There are usually insights in big amount of data, but it cannot be easily found, the data scientist need analyze the data and found the pattern behind it.

One data set can be adapted to be used across several different products. Which means one data set can be used for different purpose. For example, the Taobao transaction data can be used to analyze the sales number of different products to find what product sales well and what not.

The same data can be also analyzed to find different customer’s purchasing preferences and recommend the customized related product to them.

4

**The Model of Generating/Consuming Data has Changed, in the old model** Few companies are generating data, all others are consuming data. the few companies like newspaper agency, publishing house, the broadcast and TV station, which are the resource of data.

In new model everyone could be the resource of data, all of us are generating data, and all of us are consuming data. we publish content in blog, twitter, Facebook, WeChat and YouTube, TikTok etc. and we also read the content there.

5

**Now comes what is big data, is that when the data gets big, it is called big data? The answer is No.**

**Big data is a term that describes**

**large, hard-to-manage volumes of data – both structured and unstructured .**

**Which is not only big volume, but also is hard to manage by the traditional way we used before.**

**The traditional way we used to handle structured data, which is the data in table form with rows and columns.**

**– that inundate businesses on a day-to-day basis. ...**

**It’s what organizations do with the data that matters.**

**Big data can be analyzed for insights that lead to better decisions and strategic business moves.**

6

From the generation of electronic data to today, we experienced several stages.

In early stage, the data generated by the transaction systems as the business activity happened.

For example, when you buy something in Walmart, it would generate a sales record data in Walmart system.

Then when the social media and mobile device developed, the ordinary people can send message and publish the content in text voice and video.

So, it is the user-generated content stage.

The 3rd stage-**Perception stage**

Sensor with processing capability, automatic; (7\*24) generate data, like video camera, wearable device etc., which generate data all the time, no stop. ---This is the main reason for big data

7

This diagram shows the IT progress, since 1980s, Von Neumann invented the computer, we entered information age, the PC can generate and process the data.

And then the internet connects all the computer together, which make the data move and share among all the PC, this promote the information generation and exchanging. The information can easily transfer from one place to another end of the earth.

The popularity of mobile devices has liberated the limitation of computers on people's physical space, then people can create and publish contents wherever they are and whenever they want. Which make Big Data.

New York Times 2012 special column: The era of big data is here, decision-making relies more on data analysis. We have a lot of data from different aspects, connecting all this data and process it in the suitable way, we can extract useful information to support our decision making.

After 2010, the AI era has come and it has become the second sprout of the technological ecology after the Internet. AI give the machine some intelligence, which can analyze the data itself, which help people analyze the data and react correspondingly.

This enable us analyze more widely and deeper consequently get more insight.

Four major technologies empower are Big Data, AI, IoT, 5G.

8

In this session we learned the bigdata concepts, Thank you for your attention, if you have any question, feel free to connect me.