Chapter 9
Big Data and Service Science

Tu-Bao Ho
Japan Advanced Institute of Science and Technology, Japan

Quang-Bach Ho
Japan Advanced Institute of Science and Technology, Japan

Siriwon Taewijit
Japan Advanced Institute of Science and Technology, Japan

Hieu-Chi Dam
Japan Advanced Institute of Science and Technology, Japan

ABSTRACT

Big data is about handling huge and/or complex datasets that conventional technologies cannot handle or handle well. Big data is currently receiving tremendous attention from both industry and academia as there is much more data around us than ever before. This chapter addresses the relationship between big data and service science, especially how big data can contribute to the process of co-creation of service value. In particular, the value co-creation in terms of customer relationship management is mentioned. The chapter starts with brief descriptions of big data, machine learning and data mining methods, service science and its model of value co-creation, and then addresses the key idea of how big data can contribute to co-create service value.

BIG DATA

What Is Big Data?

Big Data is an emerging issue in information technology (IT) in the last two years. In his paper "Big Data: Defining its definition" on ZDNet (2012), Andrew Brust gave a widely accepted definition: "Big Data is about the technologies and practice of handling huge data sets that conventional database management systems cannot handle them efficiently, and sometimes cannot handle them at all. Often these data sets are fast-streaming too, meaning practitioners don't have lots of time to analyze them in a slow, deliberate manner, because the data just keeps coming."

In other words, Big Data are datasets that are very *big* and/or very *complex* that the current IT methods and tools cannot handle them well. It is worth noting that the name "Big Data" can mislead people to think of only the big size of the data, but ignore or do not know about the complexity of the data. However, the big and complex aspects of such data always go together, in which the complexity of the data is even more typical and challenging for "Big Data" than the big size. We can also see

DOI: 10.4018/978-1-4666-9562-7.ch009

this feature in the IBM definition of Big Data by four dimensions of Variety, Velocity, Volume and Veracity. The first dimension (variety) is about the cross-links between data sources which can be any type of data structured and unstructured data such as text, sensor data, audio, video, click streams, log files and more. The second dimension (velocity) is about "time-sensitive processes such as catching fraud, Big Data must be used as it streams into your enterprise in order to maximize its value to your business." The third dimension (volume) is about the ever-growing data of all types, at levels of terabytes (10^{12}), petabytes (10^{15} bytes), and even zetabytes (10^{18} bytes). The four dimension (veracity) is about truthfulness, accuracy or precision, correctness of the data. It is about uncertainty due to data inconsistency and incompleteness, ambiguities, latency, deception, or model approximations.

It has been said that Cloud Computing, Smart-Devices, and Big Data are the three emerging IT technologies that can strongly influence other disciplines. We all know that data is the source of all information that people can have. However, information usually is not ready to use but requires to be analyzed (processed) for being usable. Roughly speaking, the bigger and the more complex data the more difficult to analyze it, and therefore in many cases we still not be able to do it well (Franks, 2012).

Where Does Big Data Come From?

The phenomenon that we have much data around us than before is objective. There are Big Data in many organizations, in social, business, scientific activities that potentially contains big values. But where does Big Data come from?

The Big Data come from various sources, typically the three followings:

1. **Social Media Data:** For example each day there are 230 millions tweets on twitters in the world, or 2.7 billion comments to Facebook, and 86,400 hours of video to YouTube.
2. **Machine Data:** Such as industrial equipment, sensors and monitor machinery, web logs tracks user behavior online. For example, the Large Hadron collider at CERN (the European Organization for Nuclear Research) generates 40 terabytes of data each second.
3. **Transactional Data:** Such as Product IDs, prices, payment, manufacturer and distributor data, and much more. For example, the products sales of Amazon.com in the third Quarter of 2011 are about 10 billion USD, or the pizza chain Domino reaches 1 million customers per day.

Big Data Can Be Very Small and Not All Large Datasets Are Big

As mentioned above, the most challenge in dealing with Big Data is its complexity. For example, the power stations, planes… each has hundred thousands sensors and they make decisions to function based on analysis of sensor reading combinations. It is easy to see that such data streaming of sensors is extremely complex though their size is not as large. For example, a plane with a hundred thousand of sensors only produces 3 gigabytes data in one hour of flying (100,000 sensors x 60 minutes x 60 seconds x 8 bytes = 3Gb), but managing the combination between those sensors causes the situation so called "Big Data but not large."

In contrast, there are many cases where systems generate very large quantities of data with very simple structures or clear principles that we can call the situation as "large datasets that are not big."

Big Data and Governments

Let's take an example of the campaign of the President Barack Obama using data mining to exploit Big Data in the competition with Mitt