

Developing Data Analytics Skills in Japan: Status and Challenge

Hiroshi Maruyama, The Institute of Statistical Mathematics

Abstract:

Japan needs to develop data analytics talents quickly to catch up with the trends of using data for business decisions. The Ministry of Education, Culture, Sports, Science, and Technology started a three-year project to develop so-called data scientists. This talk reviews the findings of the first year of the project, and discuss the future challenges.

1. Introduction

Data analytics is changing the way how businesses are making decisions. Ian Ayres argues in his book *Super Crunchers* that "we are in a historic moment of horse vs. locomotive competition where intuitive and experiential expertise is losing out time and time again to number crunching.[1]" According to an article in the 2012 October issue of *Harvard Business Review*, data scientists, experts who are trained for data analytics, are "the sexiest job in the 21st century. [2]"

One of the major bottlenecks of employing data analytics is the lack of talents. A report on big data by Mckinsey Global Institute[3] predicts that by the year 2018 there will be shortage of "deep analytical talent" by between 140,000 and 190,000 in the US. Similarly, *Nikkei* Newspaper reported that Gartner estimates the shortage of data scientists in Japan as about 250,000 in the future.

We started a project titled "Data Scientist Training Network" sponsored by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in July 2013. The goal of the project is sustainable development and effective utilization of data analytics experts in Japan. Our approach is to form a network of stakeholders, that is, the universities, the industry, and the government and NPO's so that they can share a common vision and exchange ideas and best practices. This paper describes the insights obtained from the first 12 months of the project¹.

The main insight is that while in the US (especially in Silicon Valley) data scientists are often considered to be developers of systems that can take advantage of data analytics while in Japan many data analytics experts are working as service providers, that is, analyzing customer's data and providing advices based on analysis results. This implies that the skills required for data scientists in these two countries should be different. In Japan where data analytics is usually provided as a service, we need at least three types of skills; (1) analyzing data, (2) providing services, and (3) making business decisions based on data analytics.

2. Data Scientists in Silicon Valley

In the above-mentioned article in *Harvard Business Review* describes a unique attempt of

¹ For the details please visit our website at <https://datascientist.ism.ac.jp/>.

developing data scientists in Silicon Valley. This startup company called Insight Datascience Fellows Program provides six-week trainings to scientists, usually Ph. D. holders in natural sciences such as physics and biology, so that they can be ready-to-use data scientists. The training includes programming and how to use data analytics tools, and the applicants (called “fellows”) are to create so called “data product” at the end of the six-week training period. This “data product” is a working program that provides actual services on the Internet, based on data analytics. For example, “CouchTube.net,” a data product developed by one of the fellows, creates organized lists of YouTube videos of TV dramas by identifying which YouTube video corresponds to which episode of the drama series. To do this, the system analyzes various data published on the Internet, such as the database of TV dramas and the text annotated to the YouTube clips as well as the comments by other users, and scores each clip with the estimated probability of being the part of a specific episode. The fellows are required to produce such working services, from conceptualization, design, implementation, system integration to testing by themselves. Thus, computer programming skills are mandatory.

This program has started two years ago and so far there have been about 100 fellows graduated from the program, and every one of them has been hired by some Silicon Valley company such as Google and Facebook. According to the conversation that the author had with a couple of data scientists working for Facebook, Facebook has a few hundred such “data scientists” among their approximately 7,000 employees. Most of these data scientists are distributed to individual product teams, responsible for integrating data analytics algorithms into these products. In the Silicon Valley labor market, the demand for such data scientist is extremely high, and they are approached by recruiters almost every day.

O’Reilly’s survey on data scientists[4] seems to reflect the fact that the data scientists in the US has a tendency of being system developers. This survey concludes that data scientists are categorized into four groups: data businesspeople, data researchers, data creatives, and data developers. Except for the first category (“data businesspeople”), they are all focusing on programming skills.

3. Data Analytics Experts in Japan

What do data scientists in Japan look like? We conducted a study on the current situations of data scientists in Japan. This study consists of a quantitative analysis based on a survey and a qualitative analysis based on a set of interviews. Hereafter we use the term “data analytics experts” instead of “data scientists” because as we explain later the image of “data scientist” exemplified by those in Silicon Valley does not seem to apply to the situation in Japan.

The survey was done on those who passed the test of Japan Statistical Society Certificate in November, 2013. The number of valid responses were 319, in which 87.5% were male. The age distribution of the respondents is shown in Figure 1. The questionnaire includes the following questions:

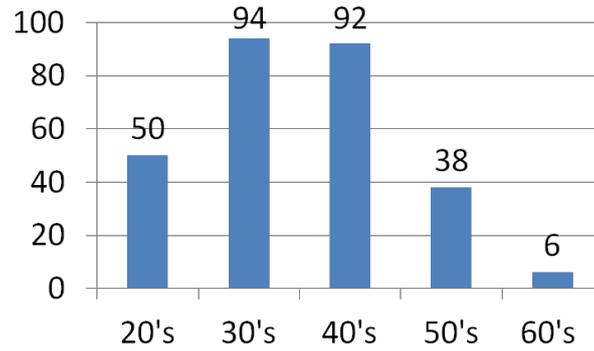


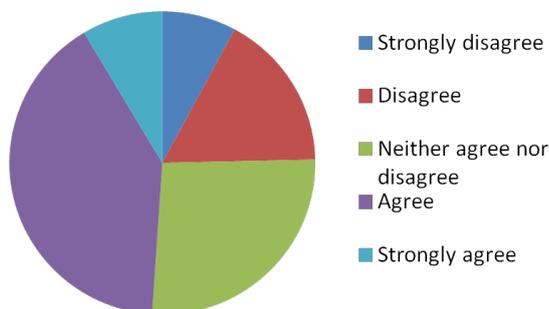
Figure 1. Age distribution of respondent

- Q 1- 3: Demography
- Q 4- 6: Industry, type of work
- Q 7-10: Data analytics at the workplace – frequency and characteristics
- Q11-18: What kinds of skills of statistics, computing, and business are needed at work and how they acquired these skills
- Q19-20: Views on career paths

From the industry point of view, the manufacturing industry and the financial industry have adopted data analytics in their business more than the other industries. On the other hand, the services industry and the public sector are lagging behind in terms of data analytics. From the skill development point of view, people who are working on data analytics generally acquired their skills by themselves or through on-the-job training. This last point suggests the lack of systematic education in this field.

Figure 2 shows the answers to the questions on career. About 50% of the respondents said that their skills are properly utilized and more than 60 % answered that they want to be a data analytics expert as their career.

Q18. Do you think your skills are effectively utilized?



Q19. Do you want to have a career as a data analysis professional?

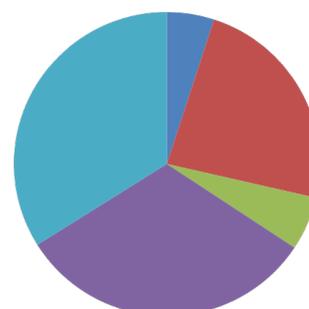


Figure 2. Views on career

For those who want to be a data analytics expert, we further asked the reason why. Figure 3 shows the results. Many view data analytics expert as a preferable career because they like data analysis, they can make use of their skills, and they think that it is important to their organization and the

society. One interesting observation is that also significant amount of responders think that data analytics expert is a fair job that is not discriminated by their age or gender.

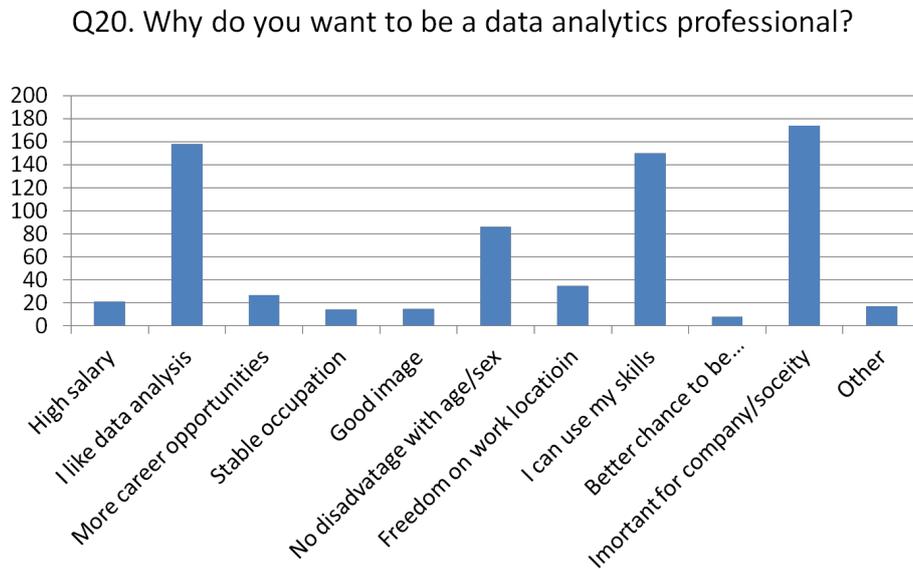


Figure 3. Reasons why they want to be a data analytics expert

As was done in O’Reilly’s report, we categorized the respondents into four groups using a clustering algorithm. With some subjective interpretation, these groups are described as follows:

1. Young people who dream to be a data analytics expert, although they do not have enough experiences yet.
2. Mid-career engineers, typically in an R&D organization in a large manufacturing enterprise. They analyze data generated within their own department. Analyzing data is part of their daily job. Their career path in the company is also well-defined.
3. Employees of small and medium businesses, typically in a service business. They chose data analytics expert as their job because of its flexibility of work style. This group has a high female ratio compared to the other groups.
4. Seasoned service professionals in an IT service company or a consulting firm. They have been providing data analytics services to their clients for a long time, and are proud of the job.

This clustering results suggests that the data analytics experts in Japan are more concerned with the type of the work (e.g., whether they analyze their own data or their client’s data) and also their long term career paths while the data scientists in the US seem to be more concerned with their skills and how to utilize them for immediate returns.

The second part of the study is a series of interviews. We selected 20 so-called “data scientists” in Japan (although the definition of “data scientist” is somewhat different from that in the US, as we

discussed), made an appointment with them, and met with them face-to-face for one-hour interview, asking their work, skills, and career paths. The interviewees are in various industries from finance, manufacturing, distribution, public sector, IT service, and consulting. Their business sizes are also diverse, from large enterprise to SMB. One of the interviewee was a freelance data analytics expert. These interviews revealed the following points:

1. The background of data analytics experts varies – from natural science, social science, mathematics, to computer science.
2. Holistic skills are required for data analytics experts. They need to understand the customer's business goals, discover data analytics problems, formulate them in mathematical forms, solve them, and present the results to the client so that the client can understand and be ready to apply them to the real business.
3. Data analytics capability is often an organizational capability. It is rare that a single individual has the holistic skills mentioned above, but as a team of experts who have different areas of expertise can perform the required data analytics tasks. In one company, every employee is required to have a basic training on data analytics.
4. The client's side literacy of data analytics is also critical. The best report of data analytics is useless if the client cannot make decisions based on it. Thus, the decision makers of the client have to have at least basic understanding of how to interpret data analytics results.

3. Data Analytics as Service

As we discussed, data analytics in Japan seem to be more service-oriented as opposed to data scientists in Silicon Valley who are more product- or system-oriented. “Services” here refer to both inter-company services (e.g., those provided by consulting firms) and intra-company services (e.g., a “Data Analytics” department in the company provides services to other departments in the same company, or other companies in the same conglomerate).

What are the skills needed for successful service engagements of data analytics? From the interviews above, we argue that there are following three types skills that are mandatory:

- (1) Deep analytics skills. These are skills for analyzing given data, such as statistics, machine learning, databases, and programming.
- (2) Service providing skills. These are skills needed to conduct service engagements, including understanding the client's business, identifying the business issues, formulating them into data analytics problems, solving and testing them, and proposing solutions. Communication skills are especially important.
- (3) Service receiving skills. These are skills needed for making business decisions based on data analytics, including the basic understanding of statistics (e.g., understanding difference between correlation and causality).

Services are said to be co-creation between the service provider and service receiver[6]. Data analytics services are no exception. The receiving side of the service has to have literacy on data analytics and be ready for making decisions. Many interviewees pointed out the importance of this. With regard to how to acquire the skills for the receiving side of service, Kijima[6] also pointed out the concept of “co-elevation,” which means that both the service provider and the service receiver learn from the engagements and acquire respective skills so that they can create larger values in the future engagements.

4. Next Steps

Our project continues by incorporating the findings described in this paper. We are currently arranging internship programs for undergraduate and graduate school students so that they can learn the skills for data analytics. At the time of writing, eleven companies offered opportunities for these students to experience data analytics in real businesses. The program offers the following three courses of internship, depending on the types of skills that students want to acquire:

- (1) Type 1: The student is assigned to a data analytics team and tasked to do deep analysis of given data. This is to develop deep analytical skills.
- (2) Type 2: The student is assigned to a service delivery team, and expected to experience a service engagement from the beginning to the end. This is to develop service providing skills.
- (3) Type 3: The student is assigned to a user department who utilizes the results of data analytics. This is to develop service receiving skills.

With this internship program we are to test our hypothesis. If it is proven, we intend to institutionalize a mechanism to develop these skills.

References

1. Ian Ayres, *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*, ISBN-13: 978-0553805406, 2010.
2. Thomas H. Davenport and D.J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review*, Oct., 2012.
3. McKinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity," http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation, 2011.
4. Harris, H.; Sean M.; Marck V. *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*. O'Reilly, 2013, 32p.
5. Kyoichi Kijima, "A Co-creation Model of the Process of Service Innovation," in *Service Systems and Systems Sciences in the 21st Century*, INCOSE International Symposium in Chicago, July 2010.