

AN EDUCATION CURRICULUM FOR DEVELOPING ANALYTICAL SYSTEM ENGINEER BEING GOOD AT STATISTICS

Naoko Sakurai

Tokyo University of Information Sciences, Japan
sakurai@rsch.tuis.ac.jp

30 years have passed since Tokyo University of Information Sciences (abbreviated into TUIS) was established in Chiba next to Tokyo, whose diversified education around informatics had developed various kinds of system engineering human resources. In spring of 2017, TUIS has started new courses, let's say Data Science Laboratory based on mathematics and statistics for SE (System Engineer), who will have ability to equip adding skills and tastes for data analysis, especially in big data. Such statistical SE is deeply demanded in modern society because everything has to be decided depending on evidence from data. TUIS's new curriculum and its future vision of Data Science Laboratory will be introduced with actual education and assessment model.

INTRODUCTION

2017 is the 30th anniversary year for TUIS in Chiba Prefecture, which has graduated system engineers in the making every year. Changing of our society as time passes, has grown information storage so large scale anytime, anywhere. SNS has been storing more message, graphics, music or kinds of words. Usual decision making process will change the position to the method depending on the results from data analysis. We can see the review by Hal Varian, who is the chief economist of Google, as follows; Data Scientist: The Sexiest Job of the 21st Century, in Harvard Business Review of October, 2012, and this also is very exciting for us of TUIS. Now various kinds of data have changed system engineering, so that it has been changeable for system engineer of its being among lots of data. For example, one system engineer of network security department, who is the graduate of TUIS, says there is so much lack of his colleague who can analyze malware because computer virus is growing in power and speed moment by moment. It is extremely necessary for working staff at computer security section to analyze stacked virus data, to build debug plan and to determine new way for avoiding enemies like virus. It is very serious problem that there is not enough data analyst in the place of making or checking system engineering section at all.

According to the paper named "Future Investment Strategy 2017" by the Japanese government office in June of 2017, the strengthening of education for data scientist against data stacked society is raised up for important and emergency task. Since the strong needs of statistics and data analysis is increasing over networking society, the faculty of data science has started in Shiga University in Shiga prefecture (Fig.1), and will start at Yokohama City University next spring. TUIS also has started data science laboratory course within mathematical informatics in Chiba prefecture in spring 2017, whose concept is along with governments' future purpose above mentioned. One of our main purposes in education is developing system engineer good at data analysis, who can extract new idea or latent discovery from complicated code. Chiba prefecture, in which TUIS is located, next to Tokyo also has many computer engineering companies, for instance, we have computer science integrated area around Makuhari, where large scaled exhibition hall named Makuhari Messe is there, to which we have feeling of symbolic place for computer integration, therefore, TUIS's mission is growing great. With this in mind, in this research the curriculum model for training data scientific system engineer will be introduced.

CURRENT DATA SCIENCE EDUCATION

Coursera has lots of data science courses, which is much popular around the world because of strong demand from the society. Coursera is one of the most famous MOOCs in English, which is a model of cooperation with industry, government and schools, offering over 2000 courses and over 180 special courses. Coursera has provided master course of data science named "Masters of Computer Science – Data Science –", students can get the master degree by studying with this course same as on campus. You can see an introductory guided sentence of this course on the top page of Coursera. (Fig.2) One model of the student's planning worksheet has introduced that three



Launch a new field! The department of Data Science is the first for faculty in Japan to meet a real society demand for rich imaginative leaders who use data efficiently.

Figure 1. Faculty of Data Science in Shiga University

requirements are needed such as breadth (16 credit hours), advanced course work (12 credit hours) and additional course work (4 credit hours). Breadth requirements contain data mining, data visualization, machine learning and cloud computing, and advanced course work has statistical analysis, information science and capstone courses. This is completely online degree served by university of Illinois and top-ranked iSchool. Many students have got the degree in three academic years.

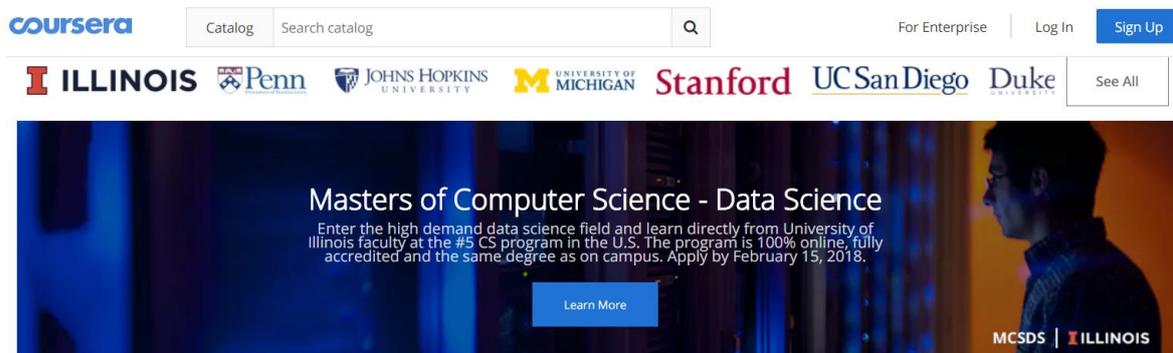


Figure 2. Accredited Special Course for Data Science

Here is another introduction of Coursera’s Data Science course in Top Specialization. Top 5 categories are related to data science, this is much exciting because of its reflection of modern data age. 5 categories are as follows;

- 1) Deep learning (5 courses)
- 2) Python for Everybody (5 courses)
- 3) Data Science (10 courses)
- 4) Applied Data Science with Python (5 courses)
- 5) Statistics with R (5 courses)

Applied Data Science with Python has 5 courses, which will start firstly “Introduction to Data Science with Python”, the second is “Applied Plotting, Charting & Data Representation in Python”, and the next is “Applied Machine Learning in Python”, the fourth is “Applied Text Mining in Python”, and the final is “Applied Social Network Analysis on Python”. Python is familiar to study data science because of its characteristics like free use, easy coding and lots of library being useful for data science (Grus, 2015). This category has exactly indicated one total image and contents for

data science, which has various kinds of definition now in complete confusion but having wide variety among sections.

On the other hand, data science education course is growing more popular on MOOC in Japanese (“JMOOC” or “gacco”) 2 years behind Coursera and others in English. There are so good data science courses, Statistics 1 to 3, Introduction to Data Science and its Exercises for business or working people, Statistical Open Data for Everybody and Introduction to Data Science and its Exercises for high school students. Some courses have direct classroom instruction, which will lead more understanding them to data insight. The establishment of gacco had been waited for a long time by the people wishing to learn statistics and data analysis because of high leveled needs from our society. Although MOOCs in English version have provided lots of data science and statistics courses, it is a little bit difficult for us in Japanese in spite of the services like as subtitles on the display. Gacco will make much growth depending on the supports from government agency and related companies.

CURRICULUM MODEL OF TUIS

Data is in existence over sorts of field in diversified condition, for example, chatting words or photo as Instagram. Every data can be analyzed to meet a new finding, which is very important for people in the world by data scientist. Consequently, it is so great to develop good data scientists by systematic education curriculum. It really is desirable for data scientists to have both sense of humanities and science, so we have to build up our education curriculum for data science along a point of view over a human scientist of data with computer. It is essential for data scientist to have power into computer science with exact algorithm for managing large scaled data, in addition to statistics and statistical data analysis. Moreover, they need the challenging spirit for unresolved problem with mathematical science and the abilities of proper visualization and communication. A system engineer also need such faculties over their work because they meet all sorts of trouble or question every day, so they have to analyze assorted data around job.

Here is an introduction of the curriculum at the laboratory of data science in TUIS which have started on spring of 2017. Figure 3 shows an excerpt of our curriculum for freshman. In this first year, students will take 3 main categories lectures, around statistics, computer science and other liberal arts. Knowledge creation includes active learning typed lecture, with which students can get their own thinking and activities. Figure 4 shows an excerpt for sophomore. They can learn not only computer science and statistics, but also mathematics and some social data analysis.

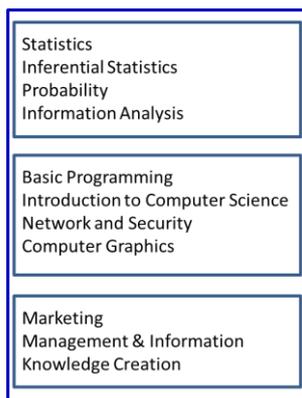


Figure 3. Freshman

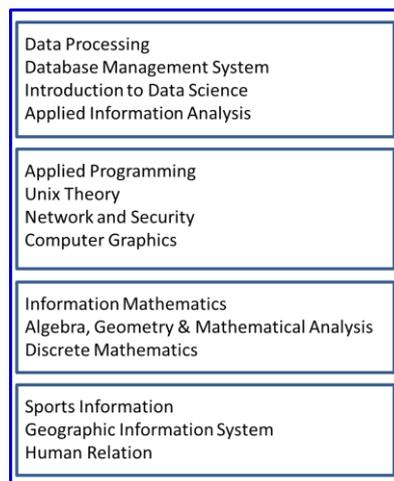


Figure4. Sophomore

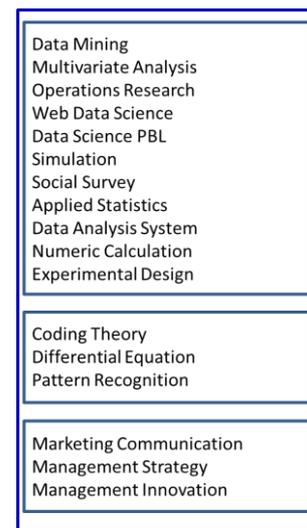


Fig.5 Junior (Third Year)

Figure 5 represents a lecture model for Junior students majoring mathematical informatics in which data science laboratory belongs. Junior students will take more statistical analysis than before, so they can be powered in analytical skills, this will lead them being superior in statistical system

engineer. In addition, since the curriculum for Junior includes around business administration lecture, students will be able to get practical business knowledge before graduation. Our curriculum has 4 main pillars such as Computer science, Statistics, Mathematics and Business Administration, in addition to these, project based training with real data will make a practical science feature.

One lecture “Web Data Science” is picked up and introduced here, which is very close to data science under networking. In this lecture, links among pages, access log or text in website are analyzed as data in data science. Students will learn statistics, mathematics, algorithm for analysis and visualization for data in website. In particular, syllabus is as follows.

- 1) Analysis and visualization of links among network as complex networking
- 2) Analysis and visualization of access log in time series or correlation
- 3) Analysis of website text (HTML source code)

There are kinds of statistical software, which are free or charged, so it is better to use along data condition and personal surrounding. Fig.6 represents an introduction page of the department.

The screenshot shows a website page for 'Mathematical Informatics'. At the top, there is a navigation bar with three main sections: 'About TUIS', 'Faculty, Course & Graduate', and 'Campus Life'. Below this, a dark blue header contains the text 'Mathematical Informatics'. Underneath, there is a section titled 'Educational Objectives and Ideal Candidates' which contains a paragraph of text: 'Mathematical informatics explores methods of comprehending and solving environmental and social problems using data analysis and mathematical models. By learning these information systematization techniques, we teach students to become data scientists and meet the needs of society by obtaining the processes and basic mathematics necessary to have a command of high level information technology for decision making.' Below this is another section titled 'Laboratories' which lists 'Mathematical Informatics Laboratory / Data Science Laboratory'.

Figure 6. Mathematical Informatics (TUIS, Japan)

SUMMARY

In this research, our new curriculum for statistical SE, who can be used to manage data with statistical mind. They will exactly be in great demand in this data evidenced age (Sakurai, 2017). The system they build up will be hoped to be humanized and developing for the future. Digital transformation is an important point of business administration, in which challenging new technology, AI, IoT or IoE, and new idea are demanded by IT customer, who can access easily to cloud service system. New value, reinforcement of basic technology and new typed analytical SE, based on Information Technology, business knowledge, mathematics and statistics are necessary for system engineering in local industries. Our Data Science Laboratory will do best training to our students to meet social requests with watching the changes in future society. Assessments will be focused along with our research up to now (Fukasawa, et. al., 2018).

REFERENCES

- Coursera: <https://www.coursera.org/>
 Harvard Business Review: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
 Grus, J. (2015). *Data Science from Scratch*. O'Reilly Media, Inc.
 Sakurai, N. (2017). An Overview of Data Science: Focusing on its Education. *Journal of Tokyo University of Information Sciences*, 21(1), 51-59.
 Fukasawa, H., Sakurai, N. & Izumi S. (2018). Statistical Exploratory Process and its assessment. *Proceedings of the Institute of Statistical Mathematics*, 66(1), in press.
 Izumi S., Sakurai, N. & Fukasawa, H. (2016). An Interactive Lecture Design of Statistics for University Student and its Assessment. *The Institute of Statistical Mathematics Joint Research Report*, 362(8), 5-10.