End of AY 2017 Report for SIP – Group 1

Project Title

ID: 17-01 Title: Knowledge and Data Management for Policy Making Towards Data Driven Society

Team

GSDM ID	Name	School	Department	Year (e.g. D1)	Leader/ member
16115	Kensaku Matsunami	Engineering	Chemical System Engineering	M2	Leader
16117	Cheng-Han Yeh	Engineering	Mechanical Engineering	D2	Leader
14126	Seonwoo KIM	Engineering	Mechanical Engineering	D2	Member
15120	Hiroshi Miyake	Frontier Science	Computational Biology and Medical Sciences	D1	Member
14108	Koki Muraoka	Engineering	Chemical System Engineering	D2	Member
13205	Anjar Dimara Sakti	Engineering	Civil Engineering	D3	Member
17109	Jie Zhu	Engineering	Chemical System Engineering	D1	Member

Objective: Explain what social/global issues that this project tried to address and why the issue is important.

In our modern society, the potential expectation of data-driven society to the extent that generating innovative businesses and creating values by both combining and exchanging data among different domains has been increased. By analyzing existing data with proper tools and good ideas, we can propose many good solutions to social problems. In AY2016, we performed a bibliometric analysis on academic papers to support decision making of the government on an evaluation of research of the different field. However, there is a significant gap between data, demands, ideas, and tools, which hinders proper unitization. Therefore, even we have solutions in hand, it is difficult to deliver our ide directly to demand. In this year's SIP, **we investigate the solution to fill the gaps between a solution to the demand**. The activity is mainly performed in two ways. The first is the analysis of public attention towards a certain event. Open data sources such as SNS data or search engine databases are analyzed to solve to sustain public attention on import events. The second is finding a proper way to deliver our solution to the organization on demand. We interview the target organization where we expect our solution can be used and study how the process of realization is processed.

Method: Explain through what kind of approaches you tried to achieve the objective.

*About the list and details of the interview, add the appendix.

In this SIP, we focused our scope on the data-assisted earthquake disaster prevention, which is increasingly demanded in Japan after the 311 earthquakes. The method was divided into two steps: (i) analysis of public attention data, and (ii) implementing the idea to the real world.

(i) Analysis of public attention data

We aimed to estimate the people's potential attention to earthquake, by developing a mathematical model on SNS data. We surveyed the online data resources which reflect the public consciousness to earthquakes. We selected Google Trends as the data we use, which provides the normalized index of the number of web searches about arbitrary keywords. Google Trends is superior regarding availability, and more importantly, the geographic information associated with each record. First, we trained each other by holding a lecture of data analysis using a programing language, R. We developed by ourselves a pipeline to recalculate a batch of Google Trends data from 2006 to 2017, and combined the data with the geological and seismological annotation of the earthquake, provided by Japan Meteorological Agency. We applied the Prophet model [1] to the time-series data to convert it to interpretable form, which consists of three factors; trend, periodicity,

and event-specific pulse. Next, we looked the persistence of the public attention deeply, by analyzing half-life time of people's attention to an earthquake after they experience it. To our surprise, people's attention to the earthquake is kept high only for one to five days, no matter how huge the disaster is. Furthermore, by analyzing the people in Tokyo, we observed the negative correlation between the half-life time of attention and the distance of epicenter. Those analyses let us observe the general tendency of the human attention to the earthquake, which is hidden behind the raw data.

(ii) Implementing the idea to the real world

We went to Tokyo Metropolitan Government (TMG) for the interview to learn how to implement the idea to the real world. The interviewee was Mr. Tooru Suda (Disaster Prevention Division, Bureau of General Affairs, Tokyo Metropolitan Government). The contents of interview were three topics: daily work routine/regular tasks, feedback of our project, and the work of disaster prevention utilizing the big data. Also, we tried to contact many academic professors: Prof. Hisashi Yoshikawa from UTokyo, Prof. Masaru Yarime from City University of Hong Kong, Dr. Kyoko Ohta from Ehime University, and Dr. Teruaki Hayashi from UTokyo to acquire the knowledge on how to apply the analysis on policy-making. We asked them what kind of stakeholders we should contact to, and how to implement outcomes of our project in the real world.

[1] Taylor, S. J., & Letham, B. (2017). Forecasting at scale. The American Statistician, (just-accepted).

Outcome: Explain what kind of results you obtained from this project and discuss how it addressed your focal social/global issues.

As a result of the first step, some hidden messages, which are rarely noticed by the authorities, were revealed, e.g., the baseline of searching frequency rose after 311 earthquakes, short half-life time regardless of the event magnitude and regional dependence. The half-life time—the time required to decay to half of the peak top—was less than five days regardless of the magnitude of the earthquake. From more detail studies focusing on Tokyo citizens reacting to the event compared with nationwide, it can be concluded that people tend to pay attention to the events at close range.

Also in the second step, current situations of the government and knowledge were founded. From the interview with Disaster Prevention Division of TMG, we found that National Institute of Information and Communications Technology (NICT) is also developing tools utilizing SNS data, which also agree with one of our project directions. While this SIP project aims at analyzing the public awareness and promoting it in long-term using the history data, the system from NICT focuses more on real-time feedback for providing a more precise solution after the disaster occurs.

Other interviews, which are held with professors, indicated that it exists the difficulties of implementation of ideas into the real world, e.g., meeting higher authorities who is related to decision-making, and the importance of the human network. Without these key factors, the barrier to implementing our ideas into the real world would be much higher.

Overall, two steps were executed in this SIP project in AY2017. In step (i), the topic of earthquakes was selected and analyzed since there are wide varieties of data and it is highly related to the public. After some trends and new findings came out during the analysis, we proceeded to step (ii), which is to implement the ideas into the real world. For this purpose, we held several interviews with the authorities related to the earthquake prevention and public policies. The feedback from the interviewee was positive with some valuable comments which can be beneficial for the project in the following fiscal year.

Budget: List the budget this project implemented. *About the details, add the appendix.

Purposes	Expense
Traveling fee for the interview (Tokyo Metropolitan Government)	JPY3,732
Total	JPY3,732