博 士 論 文 （要約）

Using Social Media Communication Data for Recovery: A Study Exploring the Possibility of Detecting Socio-Economic Activities Following a Disaster

（復興のためのソーシャルメディアデータ活用：災害後の社会経済的活動検知の可能性に着目して）

澁 谷　遊 野

The purpose of this study is to explore the possibility of using social media data for detecting socio-economic recovery activities. In the last decade, there have been intense research activities focusing on non-traditional data sources, such as social media during and after large-scale disasters. This approach, which regards people's communication on social media as a sensor of real-time situations has been widely adopted as the “people as sensor” approach. Using the “people as sensors” approach has been studied in the last decade in the context of disasters, but not in regard to the following two points. First, there is a lack of timeliness in existing socio-economic recovery indicators, such as observing changes in population and production, and conducting questionnaire surveys. Thus, using social media data has the possibility of filling the gap by its timeliness, volume, and diversity. Secondly, a large number of studies focusing on using social media data to improve disaster situation awareness investigate relatively short-term response phase, and few focus on recovery. Motivated by filling these gaps, this study particularly focuses on the relationships between people’s communication on social media, Twitter and Facebook Pages, and socio-economic recovery activities as reflected in the used-car market data and the housing market data in the case of two major disasters: The Great East Japan Earthquake and Tsunami of 2011, and Hurricane Sandy in 2012.

The approach of this thesis is designed as one of synthesis across interdisciplinary domains including the following three perspectives: disaster recovery studies, crisis informatics, and economics. More specifically, first, by focusing on market data, the author restricts the consideration of this thesis to socio-economic recovery activities of disaster-impacted communities to restart their daily lives. Secondly, to explore the possibility of improving situation awareness during recovery, this thesis focuses on the “people as sensors” approach which takes advantage of the nature of social media and analyzes relationships between people’s activities in the real world and the cyber world. Thirdly, this thesis narrows the target recovery phase and the approach to specific points. Within the recovery phase, the focus of this study is mainly on the periods from the day when a disaster occurred to the “everyday-life reentry phase,” in which disaster-impacted people reconstruct their daily lives. Among recovery research approaches, this study’s main focus is on the “mechanism approach,” specifically on exploring the possibility to improve the measurement of recovery.

The thesis is composed of five parts as follows:

In the first part, the author starts the thesis by providing the purpose and scope of this study (chapter 1) and the literature review (chapter 2). Then, in chapter 3, research questions and methodologies are explained. The main research question of this thesis is whether social media communication data can be used to detect socio-economic disaster recovery (RQ). To address this research question, the following two sub-research questions are introduced. First, as the prerequisite analyses to support the appropriateness of using both used-car market and housing market data for the primary analysis, the thesis examines whether used-car market data can be a proxy of one of the socio-economic recovery activity indicators (RQ1a) and whether housing market data can be a proxy of one of the socio-economic recovery activity indicators (RQ1b). As the second sub-research question, the thesis investigates whether there are correlations between social media communication and the socio-economic recovery activities (RQ2). For social media communication data, this study uses two different types of platforms, namely, Twitter and Facebook Pages. This study also considers both topic frequency and public sentiment in these two platforms.

In the second part, this study looks into the case of the Great East Japan Earthquake and Tsunami of 2011 and provides evidence to support the validity of using used-car market data (chapter 4) and housing market data (chapter 5) as proxies of socio-economic recovery activities by empirically showing there was excess demand

for used cars, particularly Light Motor Vehicles (Japanese car types whose engine volumes are 600cc or less), and excess demand for housing located near the building damage zones after the disaster. Interviews conducted with used-car dealers and various reports revealed that people in the disaster-affected area needed to buy used cars to resume their daily routines, such as going back to workplaces. This study, therefore, argues that the excess demand for used cars implies that affected people were willing to spend their money to purchase used cars to became active in rebuilding their daily lives, and thus the used car market data can be a proxy of socio-economic recovery activities. As to the excess demand for housing, this study argues that the housing market data can be used as a proxy of socio-economic recovery activities because previous studies have recognized that housing is one of the critical factors for life recovery and moving homes suggest that affected people took steps toward their recovery, such as leaving evacuation shelters and temporary housing.

In the third part, as the primary analysis, various types of correlations between social media communication and socio-economic recovery activities in the case of the Great East Japan Earthquake and Tsunami are analyzed. First, in chapter 6, the results about the relationship between topic frequencies on Facebook Pages and excess demand for used cars, suggest that there was excess demand for used cars in the disaster-stricken area when people on Facebook Pages communicated more about topics related to activities in recovering, emotional encouragement, and the disaster itself. On the other hand, when people communicated more about anxiety and information seeking, there might have been less demand for used cars in the disaster-stricken area. Secondly, in chapter 7, the results about the relationship between topic frequencies on Twitter and excess demand for used cars, suggest that there was excess demand for used cars when people local to the disaster-stricken area communicated more about recovery and disaster damage on Twitter. In contrast, during those periods when excess demand for used cars existed, people not local to the disaster-stricken area might have communicated more about going to and supporting the disaster-stricken area. Thirdly, in chapter 8, the analysis regarding correlations between public sentiment of social media data and excess demand for used cars suggests that there was excess demand for used cars when complex sentiment expressions, containing both positive and negative words, prevailed among people local to the disaster-impacted area on Twitter. Meanwhile, during those periods when excess demand for used cars existed, people not local to the disaster-stricken area on Twitter might have expressed fewer complex emotions. Lastly, in chapter 9, the results of the investigation about relationships between social media communication and excess demand for housing, show that there was excess demand for housing when there were more communications about opinions and means of transportation among people local to the disaster-stricken area on Twitter. On the other hand, there might have been excess demand for housing when there were fewer communications regarding nuclear-related recovery activities.

In the fourth part of this thesis, the case of Hurricane Sandy is studied with the aim of supporting the findings in the case of the Great East Japan Earthquake and Tsunami. First, the thesis shows excess demand for dry housing in New York City after Hurricane Sandy to support the appropriateness of using housing market data in New York City as a proxy of socio-economic recovery activities in chapter 10. Then, the investigation regarding correlations between public sentiment on Twitter and excess demand for dry housing is conveyed in chapter 11. The results suggest that there was excess demand for dry housing in New York City when people local to the hurricane-stricken area became more positive.

Lastly, in the fifth part of this thesis, chapter 12 summarizes the findings of this study, provides the answers to research questions and discusses contributions, limitations and topics for future study. By combining interdisciplinary approaches including several perspectives, such as an economic model and machine learning, and analyzing various aspects of social media communication data and the socio-economic recovery activities,

this thesis has addressed the main research question. In sum, the results provide quantitative evidence that social media can be used for detecting socio-economic recovery activities after the two large-scale disasters (RQ). More precisely, for RQ1a, this study identifies excess demand for used cars, particularly Light Motor Vehicles, and thus used-car market data can be used as a proxy of the socio-economic recovery activities after the Great East Japan Earthquake and Tsunami. For RQ1b, excess demand for houses located near the building damage zones after the Great East Japan Earthquake and Tsunami and excess demand for dry housing in New York City after Hurricane Sandy were empirically shown, and thus housing market data can be used as a proxy of socio-economic recovery activities. For RQ2, by illustrating various aspects of correlations between social media communication data, this study provides evidence to support the possibility of using social media data for detecting socio-economic recovery activities.

The contributions of this thesis are summarized as follows. First, the study sheds light on the "people as sensors" approach for detecting socio-economic recovery activities, which has not been thoroughly studied to date but has the potential to improve situation awareness during the recovery phase. This is because the "people as sensors" approach takes advantages of timeliness, volume, and variety of social media data, and thus has the possibility of complementing existing socio-economic recovery indicators. Secondly, the study proposes new socio-economic recovery indicators: used-car market data and housing market data. The increased demand for these two types of goods by disaster-affected people for restarting their daily lives has been portrayed by various reports from governments and the mass media after several large-scale water-related disasters. However, to the author's best knowledge, there is no academic research to explore the excess demand for these two types of goods after disasters with the aim of using them as proxies of socio-economic recovery activities. By applying an economic model to used-car market data and housing market data, this study shows that there was excess demand for used cars and excess demand for housing after large-scale water-related disasters, and thus argues that used-car and housing market data can be used as proxies of socio-economic recovery activities. Thirdly, this thesis reveals the different communication patterns between people local and not local to disaster-stricken areas. This result implies the importance of distinguishing social media data posted by people local and not-local to disaster-stricken areas for using social media during the recovery phase.