Dissertation Abstract

Personal Information Protection and Rational Utilization in Space-time-behavior Analysis Based on Big Data

Graduate School of

Natural Science & Technology

Kanazawa University

Division of Environmental Design

Student ID No: 1824052011

Name: LIN YONG

Chief Supervisor: Professor SHEN ZHENJIANG

June, 2021

Abstract

This dissertation focuses on the protection and rational utilization of personal information in space-time-behavior analysis in the era of big data. Firstly, it comparatively studies the main models and basic framework of personal information protection in major countries, analyzes the dilemma faced by the basic principle of "informed consent" for personal information protection in the era of big data, and points out the conflict between the current anonymization protection way of data sharing and the usage method of personal information in space-time-behavior analysis. Then, the theory of "Rational Expectation" is introduced, and the matrix method is applied to assess the risk of personal information utilization in specific context. The judgment standard for personal information protection and utilization based on rational expectation in space-time-behavior analysis is proposed, so as to achieve the balance of interests of relevant stakeholders and realize the protection and rational utilization of personal information in the era of big data. And then, it makes an empirical study of rational expectation rule in specific context of urban planning. Finally, taking smart campus as an example to study how to make rational utilization of personal information to balance the needs of teaching devices and services during COVID-19 pandemic, and guide healthy behaviors to ensure the quality of education and management.

Key words: personal information, Protection and Rational Utilization, space-time-behavior analysis, big data, informed consent, anonymization and de-anonymization, rational expectation, crucial public interest, interest balance

Chapter 1 Introduction

With the rapid development of modern information and communication technology, human society has entered the era of big data. Big data is widely used in all fields of social life, economic production and public management and has become an innovative element of social development. Smart city is an advanced concept of modern social development, is the product of the organic combination of people-oriented city and information city, which is to use big data, Internet of things (IoT), artificial intelligence (AI) and spatial information integration technologies to achieve the humanization, intellectualization, systematization and sustainability of urban system. The core of smart city planning and management is people-centered. It is based on the analysis framework of space-time-behavior, and makes "dynamic" analysis of urban system with the support of advanced information technology, by mining, relating, identifying, integrating all kinds of big data (especially including personal information). It combines human elements to improve the intelligence and sustainability of land space planning, natural resource utilization, urban infrastructure construction, social public service, community livelihood affairs, and so on. The smart city planning and management based on time-space-behavior and people orientation includes living space planning and life time planning based on individual behavior, and behavior guidance for urban residents facing individual behaviors. The goal and result of people-oriented smart city planning and management based on space-time-behavior is to realize the sustainable urban development and the smart daily life. That is to say, the humanization, intellectualization, systematization and sustainability of urban development, and the intelligent, healthy and low-carbon behavior mode and lifestyle of residents.

In people-oriented smart city planning and management based on space-time-behavior, spatio-temporal big data is is the basic supporting data. And the mobile phone signaling big data is the most widely used and most important spatio-temporal big data. In reality, more and more organizations are collecting data with actual-time location information to provide a variety of services, which brings convenience to people's lives. At the same time, there are also problems that personal information is illegally collected and irrational used, even leaked on a large scale, which leads to discrimination, reputation damage and personal and property injuries. In the era of big data, informed consent principle, the basic principle of personal information protection, is facing great challenges. How to use personal information

legally and rationally, not only to protect personal information security, but also to meet the needs of economic development and social public interest, has become an important issue that we are facing. We believe that it is not only necessary to protect personal information, but also to protect the rational utilization of personal information. (As shown in Figure 1)

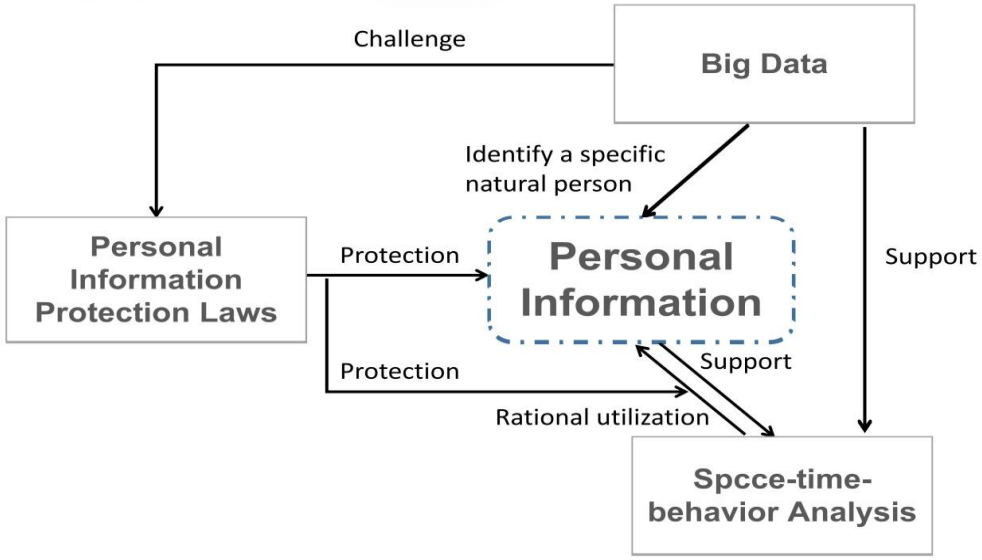

Figure 1 Logic relation diagram of background

In recent years, many researchers use planning technology and spatio-temporal big data, especially mobile phone signaling big data, to make dynamic analysis of urban spatial-temporal system, and study the characteristics of urban layout, function and human behavior. Shen and Li (2018) expounded how big data changes the ideas of planners and researchers in urban planning and management practice. By constructing the basic dynamic analysis framework of "population-time-behavior", (Zhong, Wang, et al., 2017) use mobile phone signaling big data to mine the spatio-temporal characteristics of behavior track, and identify the dynamic characteristics of individual behavior activities such as life, work and entertainment. As the basic principle of personal information protection, "informed consent" faces difficulties in the era of big data. (Lu, 2021) By using differentiated privacy protection strategy and measuring the value of privacy (Daniel and John, 2007), there is possibility and feasibility of making interdisciplinary research on personal information protection and risk management. (Pearce, 2017) Some concepts based on data context and risk management provide a new idea for the innovation in model of personal information protection. (Fan, 2016) According to the above research, we found that in the era of big data, need a new way to strengthen the personal

information protection and rational utilization in space-time-behavior analysis. In this dissertation, we take spatio-temporal big data, especially mobile phone signaling big data, as the research object. Research shows that mobile phone signaling big data is one of the most important data sources for urban planning in China. The Research Purpose is to construct effective legal rules and judgment standards to realize balance between personal information protection and rational utilization in space-time-behavior analysis. The framework of this research work is shown in Figure 2.

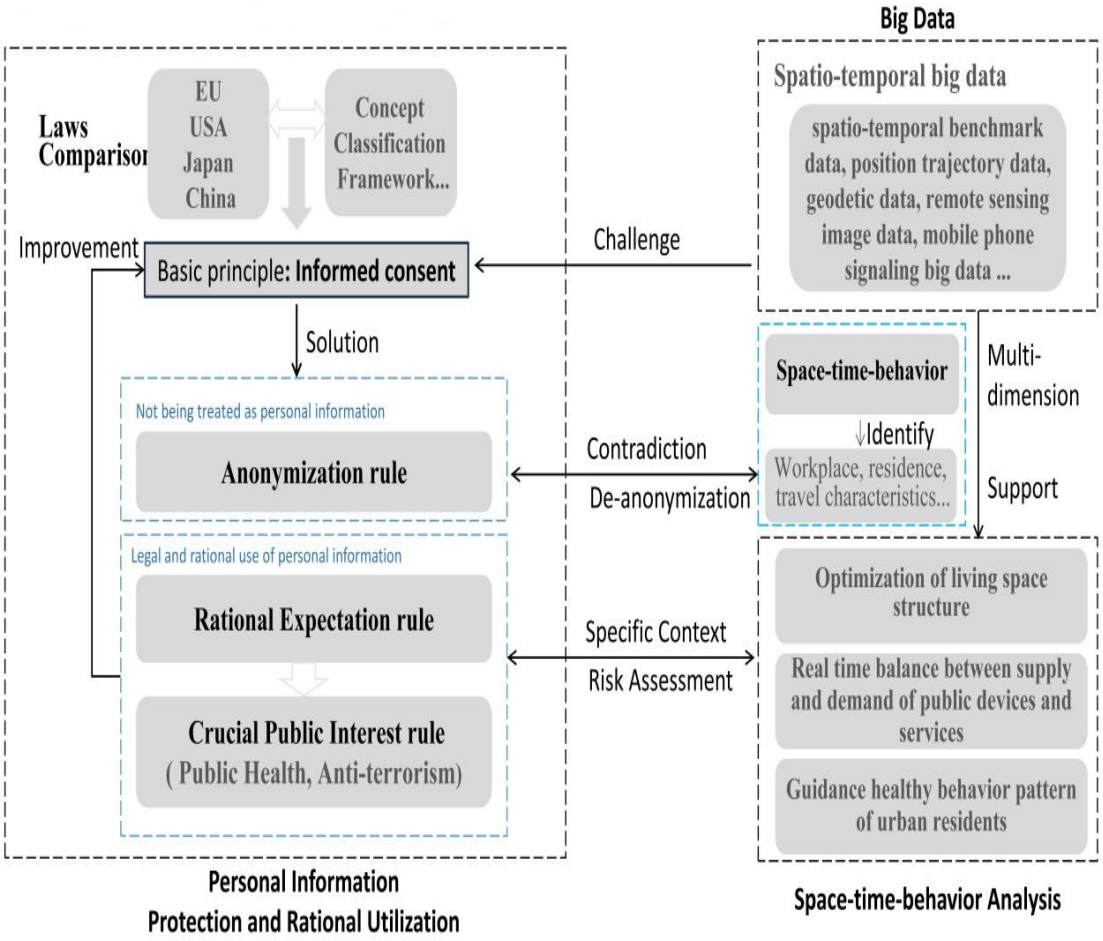

Figure 2 Framework of research work

Chapter 2 Comparative Review on Personal Information Protection Laws in Major Countries

Personal information is generally defined by "identification theory" in the international community, namely whether a specific natural person can be identified alone or in combination with other information is used as the core criterion for determining personal information. Identifiability, idiosyncrasy, correlativity are three

important characteristics of personal information. Personal information is divided into sensitive personal information and general personal information as per the sensitive attributes of the information. Sensitive personal information is relevant to one's political opinions, religious thoughts and beliefs, union membership, race and ethnicity, place of birth and residence, trajectory, health care, sexual life, criminal record, and so on.

Personal information protection legislation, United States and European Union as typical representatives. United States is the mode combining scattered legislation and industrial self-discipline, and it takes privacy right and freedom as the basis of constitution and administrative law. European Union mode can also be called as the unified legislation mode, and namely, formulating a comprehensive protection act on personal information to stipulate the collection, processing and use of personal information, this act is uniformly suitable for public departments and non-public departments, and moreover, it sets up a general supervision department for centralized supervisory. European Union pays more attention to the basic rights and interests of data, while United States emphasizes the free market model based on strong supervision. On the surface, the two models of the European Union and the United States are incompatible with each other, but in fact they share the same goal. They both seek a balance between data rights protection and data free flow. The EU model is biased towards the "data rights protection" side, which aims to create the basic data rights of citizens; However, the U.S. model is inclined to the "free flow of data" side, which aims at the development of digital economy. We believe that the protection of personal information rights is conducive to the development of digital economy, and the key is to pay attention to the balance between data rights protection and data circulation. In the era of big data, the balance between personal information protection and the development of digital economy should be a trend of data legislation.

Informed consent is the basic principle of personal information protection. It refers to the collection, processing and utilization of personal information, which must be based on the premise of fully informing the information subject and with his own (or guardian) true consent. The principle of informed consent includes two meanings: one is "informed", that is, the information subject's understanding of the personal information that the information collector will or has collected and used; The second is "consent", that is, the information subject's permission to the information collector's behavior. The principle of informed consent aims to protect the information

self-determination of individuals. However, in the era of big data, information overload, status asymmetry, data explosion and rapid transmission are challenges to the principle of informed consent. In the era of big data, some special rules are also applied to the collection, processing and sharing of personal information, to over the shortage of the principle of informed consent in practice. Generally speaking, there are two main ways. One is to de-identify personal information, so that it is no longer treated as personal information and the object of legal protection, such as Anonymization rule. The other is to make the use of personal information is legitimate, rational, necessary and risk controllable, such as rational expectation rule, crucial public interest rule, and so on. (As shown in Figure 3)

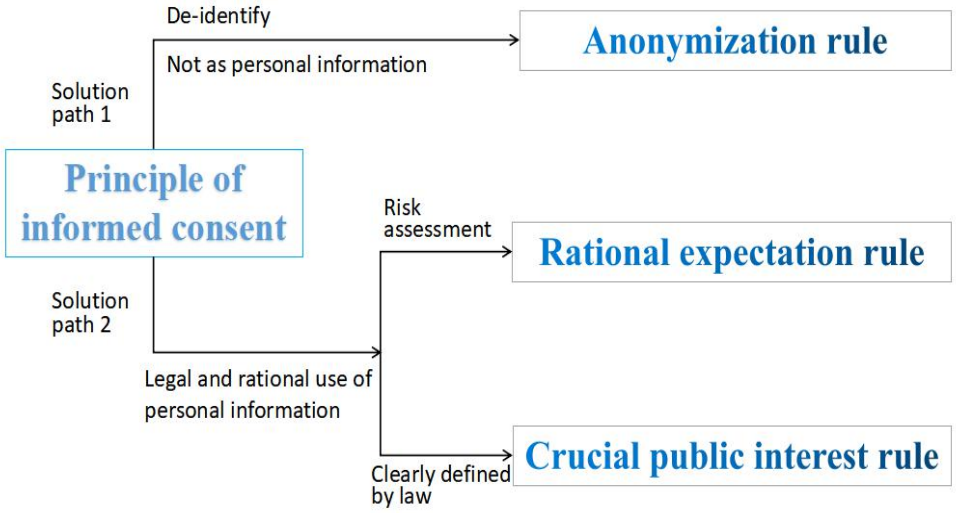

Figure 3 Logic relation diagram of background

Chapter 3 Contradiction in Space-time-behavior Analysis Based on Big Data: Anonymization VS De-anonymization

This chapter takes mobile phone signaling big data as the research object. As a control command in a mobile telecommunication system, mobile phone signaling is used for its prime technical purpose to control the link of channel and transfer the management information of communication network to uphold the normal operation of the telecommunication system. With the quick advancement of big data technologies, mobile phone signaling big data has the teatures of whole sample, low cost, high volume, high velocity and well timed, visually provides the space position, space pattern and time stamp of information. It is used extensively in space-time-behavior analysis, and its technical and economical value has been enormously broadened. Data sharing is one of the main ways to use personal

information in the era of big data. The planning based on space-time-behavior often involves data sharing between different information controllers and reusers. Before data sharing, personal information needs to be processed anonymously. Anonymization refers to the de-identification of personal information. General speaking, if the information involving individuals is processed anonymously, the contact between the information and the information subject is blocked and cannot be re identified under the existing technical conditions, it is no longer considered personal information.

The mobile phone signaling big data not only has time and space dimensions, but also has significant human behavior attributes. Even if the mobile phone signaling big data has been processed anonymously, it will inevitably show some specific location attribute information of users. The anonymous track information can be matched to the corresponding geographical space , so as to mark the active location information of the information subject in a specific period of time. It can easily identify the specific location information such as the work-place and residence of users, and even give users portrait. De-anonymization is a data-mining tactic in essence, which is a technology to re-identify the identity of information subject from anonymous dataset by technical means. The most common usage is using specific time-location method to identify the positions, then link and match the known datasets from other sources with overlapping identifier attributes to identify the information subject. Patent inquiry shows that the mobile phone signaling big data is easy to be de-anonymized in space-time-behavior analysis. It is found that only a limited number of attributes are needed to generally re-identify personal information subject with high reliability, even though the anonymous datasets are imperfect. Anonymization rule are not applicable to the sharing of mobile phone signaling big data in the space-time-behavior analysis. Mobile phone signaling big data can reflect the user's track, which belongs to sensitive personal information. Once leaked or abused, it is easy to infringe personal privacy of information subject. Therefore, only using current anonymization means to share the mobile phone signaling big data are not enough to protect the security of personal information in space-time-behavior analysis, and sharing the mobile phone signaling big data should follow the basic principle of explicit informed consent or other appropriate rules.

Chapter 4 Personal Information Protection and Interest Balance Based on Rational Expectation

Personal information has many values involving with personality dignity and freedom, economic use, and public management. Among them, personality dignity and freedom is the core value. Meanwhile, the stakeholders relevant to personal information have become more and more diverse, leading to increasingly urgent demand for sharing and using personal information. With the great improvements in the processing efficiency and transmission rate of personal information, it has become much easier to share personal information, which makes the application of the principle of informed consent more difficult. In this circumstance, theory of "Rational Expectation" becomes a new option of personal information protection in the era of big data. That is, in a specific context, the information processing should meet the expectations of the information subject, and the specific information processing should match the specific context. Personal information collected in a specific context shall not be processed beyond that context. The key to protecting personal information is to ensure the "Contextual Integrity" of the flow and utilization of personal information. How to judge whether the personal information processing context exceeds the rational expectation of the information subject, we need to introduce the risk management methods to evaluate the risk degree of the application context, so as to ensure the security of personal information processing behavior.

To calculate the risk, it is necessary to determine the factors that affect the risk, the interaction between the factors and the specific calculation method. In the risk assessment of personal information sharing, the impact factors involved in the calculation of risk value are generally personal information subject interest, threat and vulnerability, and the interaction between the factors is shown in Figure 4. Firstly, the interest, threat and vulnerability of personal information are identified (according to the existing security measures). Then, through the impact of possible threats and the vulnerability of existing security measures, the possibility of damage to the rights and interests of personal information subject is determined. And through the impact of threats and the importance of information subject interests to determine the extent of damage to the rights and interests of personal information subject. The damage to the rights and interests of the personal information subject generally includes the restriction of the information subject's independent decision-making power, the violation of humanity dignity (such as causing discriminatory treatment), and the

injury of personal and property, etc. The basic model of risk calculation is as follows:

R (I, T, V) =R (P(T, V), S (T, I))

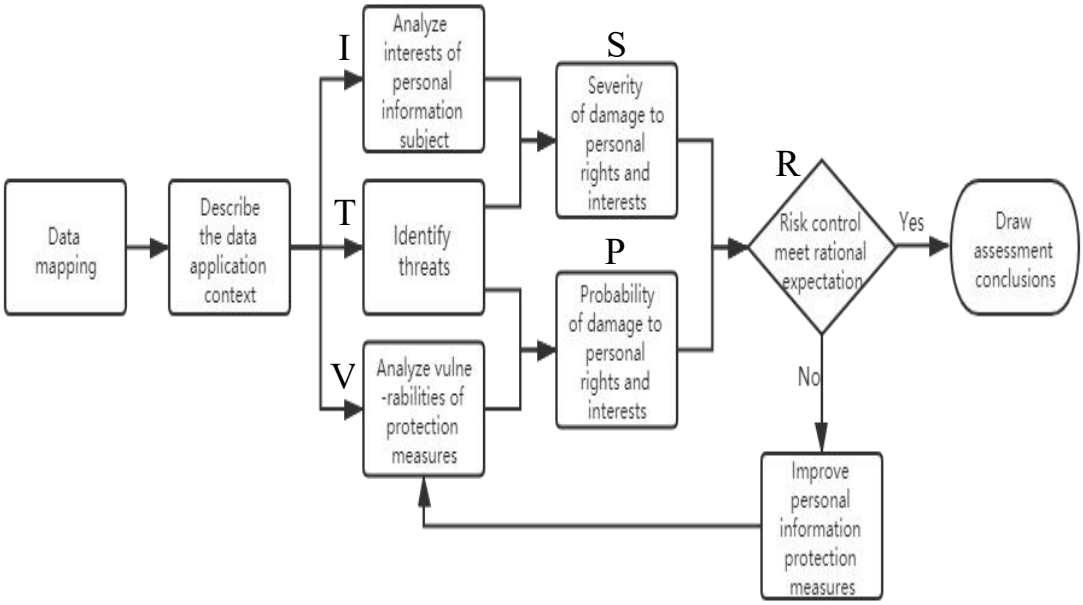

Figure 4 General Flow of Risk Assessment on the Sharing of Personal Information

By assessing the risk of personal information sharing with matrix method in application contexts, it discusses the criteria of risk control under rational expectation rule. (As shown in Figure 5) If the risk assessed is at the level of low risk, the sharing and use of personal information in this context complies with the rational expectation rule; If the risk assessed is at the level of medium risk, it is necessary to take measures timely and actively to reduce the risk and reassess the risk; If the risk assessed is at the level of high risk, the rational expectation rule is not applicable, the personal information controller should significantly inform the information subject and obtain consent before sharing the personal information. If there are multiple risk points in the application context, when and only when each risk level judged must be low risk, the rational expectation rule can be applied. Based on the rational expectation rule, we can achieve the balance of interests among personal information protection, digital economic development and public interest maintenance, so as to coordinate the promotion of digital innovation, economic development and social progress, and realize the unity of effective protection and rational use of personal information.

Risk level		Possibility of damage		
		Less possible	Rationally possible	Considerably possible
Severity of damage	Severe harm	Low risk	High risk	High risk
	Moderate harm	Low risk	Medium risk	High risk
	Minimal impact	Low risk	Low risk	Low risk

Figure 5 Matrix for Accessing the Risk Level of Sharing Personal Information

Chapter 5 Rational Utilization of Personal Information in Smart Campus During COVID-19 Pandemic

Smart campus comprehensively uses emerging information technologies such as big data, Internet of Things, intelligent perception, and knowledge management to fully perceive the physical environment of the campus, intelligently identify the learning, work situation and individual characteristics of the teacher and student groups, and integrate the school physics Space and digital space are organically connected to create an intelligent and open education and teaching environment and a convenient and comfortable living environment for teachers and students, change the way teachers and students interact with school resources and environment, and realize personalized and innovative services based on people. During the period of COVID-19 pandemic, universities may use personal information rationally to balance the needs of teaching devices and services, guide students' good learning behavior and healthy life behavior, ensure education and management quality. Usually, universities will postpone the opening date of campus and conduct online education when the pandemic is severe, will increase efficiency in campus management and start offline education when the pandemic is slow down. For the need of crucial public interest such as COVID-19 pandemic prevention, universities can collect and use necessary personal information of students, such as family background, study and rest time,

travel information, health information and so on. But it involves personal privacy, it is essential to improve the protection of these personal information in both online and offline education. In smart campus, based on the analysis of space-time-behavior, rational utilization of personal information and the combination of online and offline education will greatly help to improve the quality of education and management.

Chapter 6 Conclusions

Conclusions: This dissertation holds that it is not only necessary to protect personal information, but also to protect the rational utilization of personal information, to achieve the balance among the protection of personal information, the development of digital economy and the needs of public interests. Informed consent is the basic principle of personal information protection, but it can not fully apply to the protection and rational utilization of personal information in the era of big data. Anonymization rule is not applicable in space-time-behavior analysis and management, while rational expectation rule based on risk assessment in specific context is the improvement and supplement of the principle of informed consent. Crucial public interests can be regarded as a collection of personal interests. Crucial public interest rule is an extreme case of rational expectation rule and can be used in the crisis or emergency of public. based on rational expectation rule and crucial public interest rule, it helps to achieve a balance between the protection and rational utilization of personal information.

Legislative proposals: First, authorizes independent third parties in the form of legislation to conduct risk assessments on the application contexts of personal information; Second, allows the use of personal information based on public interest (e.g. COVID-19 pandemic prevention) through legislation; Third, plays the roles of industry self-discipline fully to achieve comprehensive protection.

Further research: This dissertation mainly from the perspective of law to analyze the protection and rational use of personal information. However, the understanding of the value of personal information is also influenced by cultural traditions, religious beliefs and social values. In the future research, we will consider more factors outside the law to build a more perfect the context-aware system. In addition, big data deep mining may generate sensitive information and affect the protection and rational utilization of personal information, which must be fully considered in future research.

令和３年８月２日

学位論文審査報告書（甲）

１．学位論文題目（外国語の場合は和訳を付けること。）

Personal Information Protection and Rational Utilization in Space-time-behavior Analysis Based on Big Data

（和訳）：ビッグデータに基づく時空間行動分析における個人情報保護と合理的利用

２．論文提出者　(1) 所　　属　環境デザイン学　専攻

(2) 氏（ふり）名（がな）　林（りん）　勇（ゆう）

３．審査結果の要旨（600～650字）

林氏の学位論文は、都市計画へ適用するため、ビッグデータに含まれる個人情報の合理的利用に関する内容である。ビッグデータを関連法制度に基づいて都市計画へ活用する際、個人情報の利用実態と問題点を明らかにすることは重要な課題である。

既存研究では、ビッグデータを活用して計画分野への適用研究が多くみられるが、計画分野への適用のため、個人情報保護の仕組みに関する研究は新規性がある。本研究は、欧米、日本の関連法制度を比較して個人情報保護の仕組みを解明し、計画分野への適用のため、個人情報を合理的に利用する必要性と公共利益を考慮する必要性の観点からビッグデータの利用原則を論じた。まず、欧米、日本の関連法制度を比較したことにより、個人情報の利用は個人の同意を前提とした本人通知制度を原則とする個人情報保護の法的仕組みを明らかにした。そして、都市構造の時空間的解析に用いられる携帯電話データの扱いについて、個人情報が削除されても、個人の移動特性を特定できる多くの特許が存在する実態を考察し、個人情報を合理的に利用するため、原則的にリスク判定手法の必要性を論じた。なお、感染状況を抑えるため、コロナウイルスの影響を受けたオンライン教育における学生の個人情報の活用に関連して事例研究を行い、公共利益を考慮しつつ個人情報保護とのバランスを考慮することが原則として必要であることを考察した。

林氏は、在学中において、学位論文の参考論文として、査読論文２編（うちESCI2編）、国際会議プロシーディング２編を公表した。なお、副論文１編がある。本審査委員会は、林氏が学位論文審査基準を満たし、必要な研究成果を挙げており、博士（学術）に値すると判定した。

４．審査結果　(1) 判　　定（いずれかに○印）　（合　格）・　不合格

(2) 授与学位　　博　士（学　術）