

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 23 日現在

機関番号：12601

研究種目：基盤研究(A) (一般)

研究期間：2014～2016

課題番号：26249074

研究課題名(和文) 携帯電話を利用した国土スケールの人口分布や移動の動態解析とその応用

研究課題名(英文) National scale analysis on population distribution and movement using mobile phone data

研究代表者

柴崎 亮介 (Shibasaki, Ryosuke)

東京大学・空間情報科学研究センター・教授

研究者番号：70206126

交付決定額(研究期間全体)：(直接経費) 31,900,000円

研究成果の概要(和文)：防災計画・災害対応、交通インフラ整備計画立案、環境管理などを効果的に進める上で、人々の分布やその移動状況に関する正確なデータは不可欠である。特に携帯電話データからはほぼリアルタイムに人口分布変化や人々の移動状況を把握でき、長期に継続観測できる可能性があることを考えると、人々の分布変化、移動状況等を携帯電話データから定量的に推定する研究は発展の余地が大きく、同時に社会に大きな貢献・インパクトを与える成果が得られると期待できる。
本研究は、世界展開を念頭に、携帯電話データを利用した国土スケールの人口分布や移動の動態把握手法を開発し、世界で初めてバングラデシュ全土の人流等を明らかにした。

研究成果の概要(英文)：Population distribution and its movement is a key information for better informed decision-making in disaster risk management, transportation planning and environmental conservation. Using mobile phone systems, we could estimate the distribution and movement of population. The estimation with mobile phone data will have a significant impact to and make contributions to society.
This study, keeping global scale applications in mind, develops a series of methodologies of estimating the distribution and movement of population at national scale using mobile phone data or CDR (Call Detail Record) data. The methodologies include a spatio-temporal interpolation method of CDR data, a method of estimating demographic attributes to create Dynamic Census data. The methodologies are implemented as a system that could be easily transplanted. Finally the system are applied to mobile phone data of Bangladesh (27 million subscribers). Population distribution and movement were successfully estimated.

研究分野：空間情報科学

キーワード：携帯電話データ データ解析 人口分布 人流 バングラデシュ 災害対応

1. 研究開始当初の背景

携帯電話から得られる GPS 情報や基地局との通信情報を利用して多数の携帯電話利用者の位置を追跡し、交通管理や災害対応に利用しようとする研究は(Ratti, 2006)が最初である。これ以前にも少数の被験者に対して携帯電話を支給しその行動を追跡する研究は行われていたが、せいぜい数百人~千人単位であり、(Ratti, 2006)によって初めて百万人単位の人々の分布変化が把握された。しかしこれらの研究は携帯電話利用者だけを対象に利用者分布の変化を視覚化し、携帯電話データの利用可能性を示したにすぎず、例えば利用者データから実人口分布を推定しその精度を評価するなど定量的な検討はきわめて遅れている。

防災計画・災害対応、交通マネジメントやインフラ整備計画立案、環境管理などを効果的に進める上で、人々の分布やその移動状況に関する正確なデータは不可欠である。特に携帯電話データからはほぼリアルタイムに人口分布変化や人々の移動状況を把握でき、今後は長期に継続観測できる可能性があることを考えると、人々の分布変化、移動状況等を携帯電話データから定量的に推定する研究は発展の余地が大きく、同時に社会に大きな貢献・インパクトを与える成果が得られると期待できる。

しかしながら世界的には GPS データが 24 時間得られる環境はきわめて希である。人々の移動等の情報が不足し携帯電話データ解析の恩恵が特に大きい途上国・中進国では、通話の度に発生する通話記録データ (CDR : Call Detail Record) しか得られない。CDR データは計測の頻度が低く、かつ位置情報は通話・通信時に使用した基地局の位置で代替的に記述されるため、精度も低い。都市部で 100~数 100 メートル、それ以外では数キロメートルになる。しかしこのデータからでも人口分布や移動動態を精度よく推定できれば、この技術を世界中に展開できる。

2. 研究の目的

世界展開を念頭に、携帯電話データを利用した国土スケールの人口分布や移動の動態把握・解析手法を開発する。バングラデシュを中心に、大量の携帯電話データを取得し、同手法を適用してその有効性を実証的に明らかにする。明らかとなった動態データを利用して深刻な災害リスク・環境リスクの下で人々が行う避難やリスク回避行動を分析できる可能性があることを示す。

3. 研究の方法

研究の方法は以下の段階からなっている。なお、国際的チームで研究を進め、また海外への情報発信を重要視したため、報告の一部が英語となっていることをご容赦いただきたい。

1) 携帯電話からのデータを網羅的に収集・

解析し、国土スケールでの人口分布の変化や移動等の動態情報を迅速に把握・解析する手法を開発した。

2) バングラデシュ全土を対象に CDR データの分析を実現し、動態把握・解析システムを適用し、人口分布、移動動態情報等を推定した。

3) CDR データから構築された人口分布や移動の動態情報を用いて、ダッカ広域圏や洪水常襲地域等の人々の移動動態等の分析を行い、交通インフラ調査、公衆衛生政策、災害対応策などの改善に適用できることを示した。

4. 研究成果

4.1 Development of CDR analysis system.

We have developed CDR Data Storage system for archiving large-scale CDR data. Currently there is lack of coordination and communication tools to provide the required information on dynamic movement of people which could be important for many cases such as earthquake and disease outbreaks. A promising secure CDR Data repository lies in a comprehensive solution. The CDR Data repository solution consists of the total solution from CDR Data anonymization until the data archiving for future use. It intends to will provide the necessary tools to address the issues on historical data storage of the CDR anonymized data for the beneficiary countries to be used in case of potential future emergencies. The Secure CDR Data Repository Solution including the web-application platform, service agent and Hadoop system for large-scale archiving.

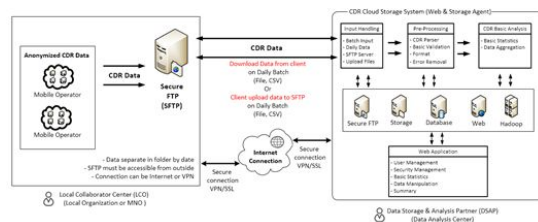


Fig.1 Configuration of the CDR analysis system

We have collaborated with BUET (Bangladesh University of Engineering and Technology), Bangladesh for analysis of CDR data in Bangladesh. We obtained 45 days of CDR data with three periods: 2015 December 1-7, 2016 May 1-31 and 2016 Oct 1-7. We conducted basic statistics on the dataset and do further analysis in three periods for population estimation as well as Origin-Destination (OD) analysis. We developed algorithms on parallel system

(Hadoop) to process large-scale. The algorithms consisted of home estimation, population estimation, Call-To-Call Movement with home padding and OD estimation.

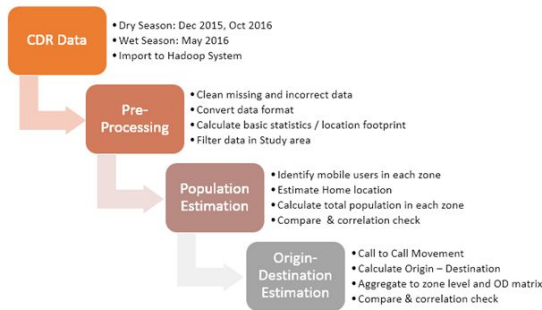


Fig.2 Steps of CDR data analysis

Since the system is designed for easy transplantation and generic CDR processing and analysis, we could easily setup and prepare CDR Data Analysis system in Sri Lanka in collaboration with LIRNEasia. One month of CDR data of Sri Lanka have been retrieved and processed for basic statistics. The CDR data included Voice, GPRS and SMS. In addition, it also provided Money Refill data. For Voice Data, caller and callee are included. The further analysis will be conducted on dynamic census.

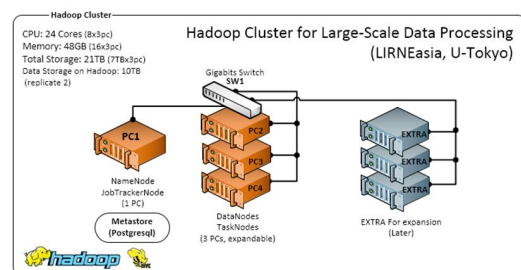


Fig.3 Hadoop cluster for large scale CDR data analysis

4.2 Development of “ Dynamic Census ” with CDR analysis

While the analysis of CDRs seem to be prominent as the means of understanding human mobility dynamics, it has been increasingly pointed out that there are constraints for utilizing such data for societal issues. The constraints are summarized as below;

1) Sparseness

CDRs provide a partial view of human mobility, which differs from a full picture of human mobility. Because CDRs are updated only when the mobile phone is used, the number of records per person per day is limited.

2) Representativeness

The population of CDRs is only mobile phone users. Thus, there can be discrepancies in human mobility dynamics between mobile users and actual living populations, which consist of mobile users and non-mobile users, if the characteristics of the two population groups are different. When we interpret the analysis result of CDRs to address societal issues, we need to be aware what population is under the study.

3) Anonymity
 CDRs are generally anonymized to protect the privacy of mobile users. Instead of the attribute information of mobile users, a random code is assigned to each person, which still allows us to trace the mobility of each person for a given period. This feature makes it difficult to know who are in the data except the fact that they are mobile users.

Increasing bodies of study analyzing CDRs have proposed to utilize mobile data to transportation study and urban planning. In fact, mobile phone data can provide the quantitative aspect of human mobility dynamics such as the volume of human mobility. On the other hand, the data alone do not indicate much about their qualitative aspects of the population, and thereby the research on the characteristics of the population in CDRs is limited so far. In this study, a novel approach is proposed to resolve the three constraints of CDRs by proposing a new dataset, Dynamic Census, by utilizing CDRs of Dhaka from one of leading telecommunications companies in Bangladesh (called the operator). The uniqueness of this study is that hidden properties of CDRs are revealed and estimated, analyzing CDRs in combination with field survey data. To understand the population structure and personal attributes, including calling behavior along with their activity, two field surveys (SPACE 2013 and SPACE 2014 – the Survey on Patterns of Activity for Comprehensive Explorations of Mobile Phone Users in Dhaka) were conducted.

Basic features of Dynamic Census are described as below;

1) Interpolated spatio-temporally

The path of the mobile user can be determined by a pair of two consecutive records in CDRs. The pair of points, consisting of time and location information, is potentially neither the time nor location of departure/arrival for a person. In Dynamic Census, blank time

bands are interpolated based on the estimated timing of location changes of the mobile users, and then route of their movement is predicted based on the existing road network. It enables us to obtain better spatio-temporal distribution estimates for mobile users.

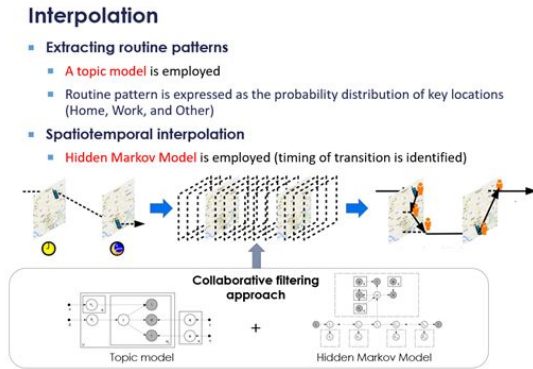


Fig.4 Spatio-temporal interpolation

2) Represent the living population for a given area

The demographic structure of Dynamic Census is adjusted to that of the living population. It enables us to discuss the human mobility dynamics not only of the mobile user but also the living population, presenting for given areas at given time.

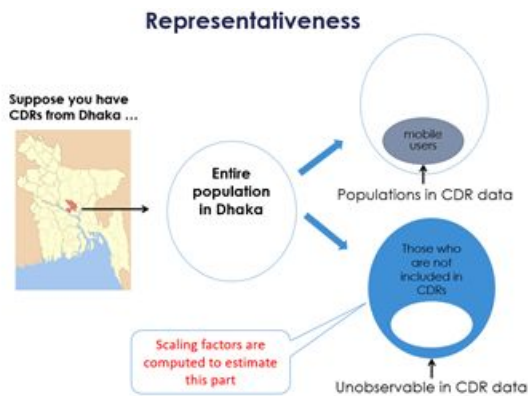


Fig.5 Enhancing representativeness of CDR data

3) Labeled with demographic attribute
Dynamic Census is labeled with the personal attribute information, which is estimated through this study. It enables

Demographic attribute estimation

- Demographic attribute estimation
 - Random Forest is employed
 - 58 user data from volunteers are used as training data
- Field survey
 - Interviewed all members in mobile phone user households
 - Two-stage stratified sampling (Land use & income level)
 - Interview survey on demographic attribute, travel-activity, and mobile phone use



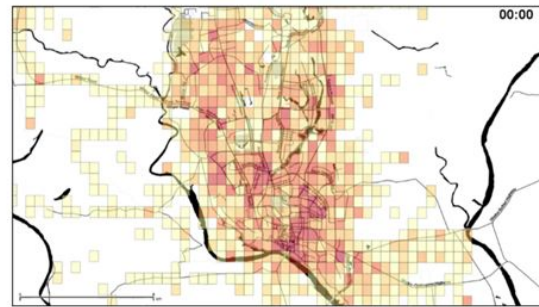
Interview at a slum household

| Estimation results | | | |
|--------------------|----------|-----------|--------|
| Class | Accuracy | Precision | Recall |
| Working male | 0.79 | 0.63 | 0.70 |
| Housewife | 0.67 | 0.47 | 0.82 |
| Student | 0.89 | 0.40 | 0.22 |
| Other | 0.63 | 0.20 | 0.07 |

Fig.6 Estimating demographic attributes by combining CDR data and field survey data

us to specify the mobility of specific population groups under study. For instance, Dynamic Census can facilitate the intervention under a disastrous event by providing the distribution of people, who are vulnerable to disaster such as small children.

Finally, dynamic census was developed



Working males' 24 hour population distribution in Dhaka

Fig.7 Example of dynamic census estimated with CDR data analysis and field survey.

4.3 বাংলাদেশ全土を対象としたCDR分析による人口分布・動態の推定

4.1 および 4.2 で開発した手法を用いて、 বাংলাদেশ全土を対象に人口分布、移動動態の推定を行った。これは世界初の試みである。なお、人口属性の推定については治安上等の理由もありダッカ都市圏以外の現地調査を実施できなかったため、ダッカ都市圏 (図7) 以外には行っていない。

図8は収集された CDR データの一日あたりの平均ユニーク利用者数である。およそ、25 百万人から 27 百万人が記録されている。

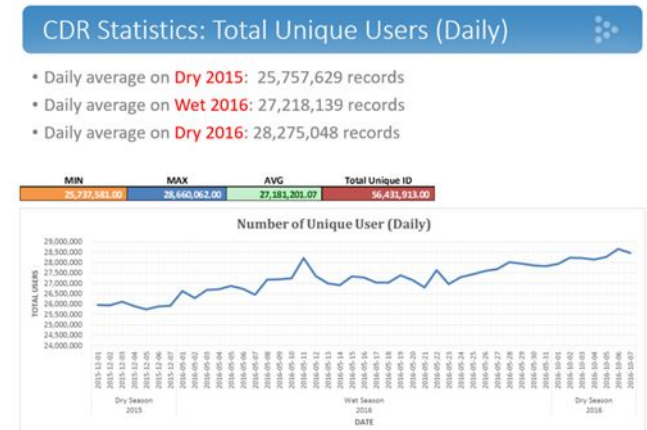


図8 収集された CDR データの一日あたりの平均ユニーク利用者数

図9は CDR データがカバーしている地理的範囲を示している。 বাংলাদেশ全土がほぼまんべんなくカバーされていることがわかる。ダッカ都市圏が特に密度が高い。

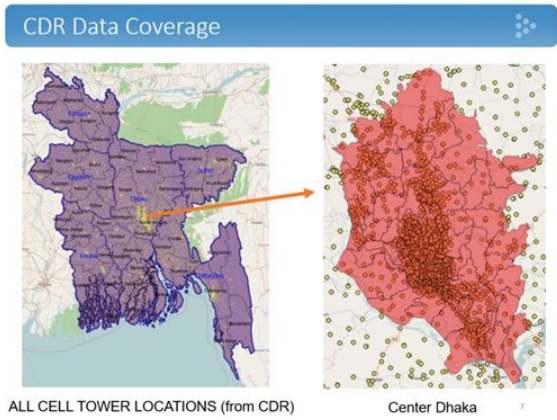


図 9 CDR データによる地理的なカバー範囲

図 10 と 11 はそれぞれ CDR データから推定された人口分布図 (Traffic Analysis Zone (TAZ) 単位) と、2014 年のセンサスデータの比較結果を示している。人口分布はほぼ高い相関が得られている。

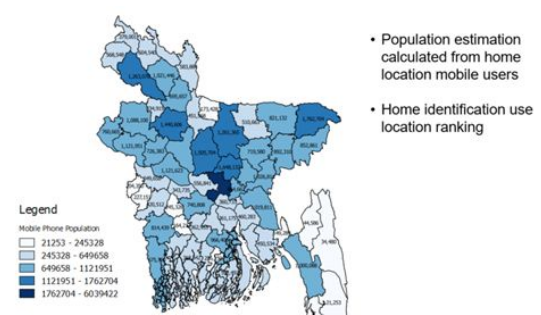


図 10 CDR データから推定された人口分布

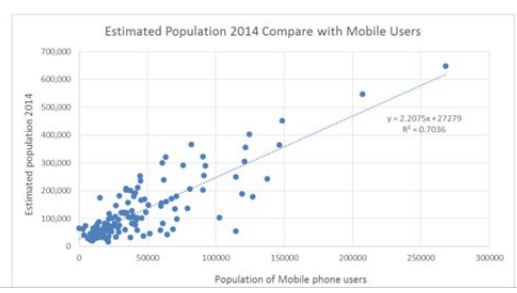


図 11 CDR からの推定人口(縦軸)と 2014 年センサスとの比較 (TAZ 単位)

図 12 と 13 は、それぞれ乾期、雨期の人口分布をグリッドベースで示したものである。CDR データ分析からは居住地が明らかになるためこうした分析が可能になる。また、図 14 と 15 は、乾期・雨期における人々の毎日の移動 (通勤等) を時間ごとに示している。これも、絶えず位置情報を得ることのできる CDR データならではの情報であると言える。

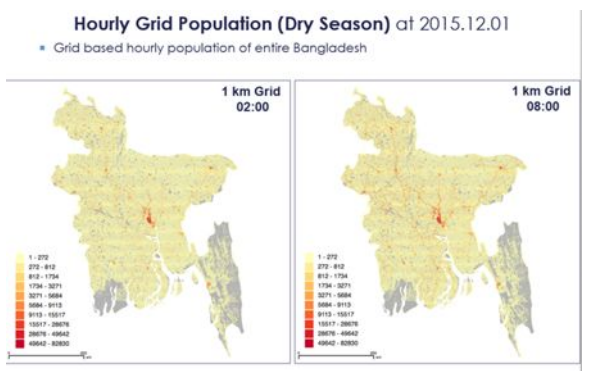


図 12 乾期 (12 月) における人口分布

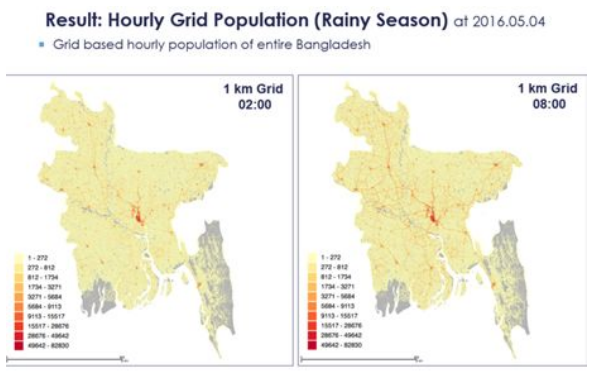


図 13 雨期 (5 月) における人口分布

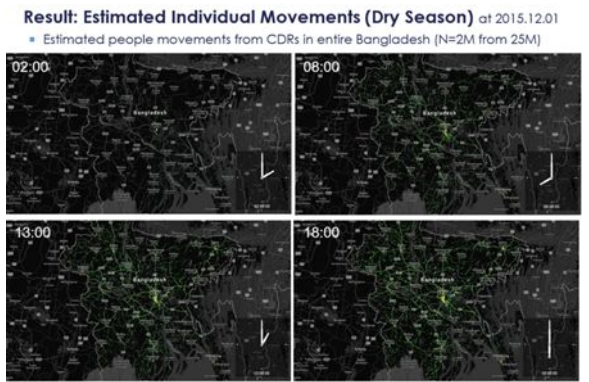


図 14 乾期 (12 月) における人口流動(動画)

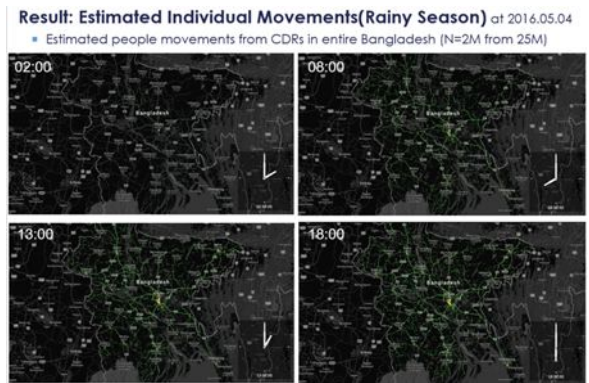


図 15 雨期 (5 月) における人口流動(動画のスナップショット)

4.4 推定された人口分布・動態データの苦痛や災害対応への適用可能性の検討

適用可能性については、それぞれ JICA によるダッカ都市圏交通調査および BUET(バングルラデシュ工科大学)との共同研究により行った。前者については訪問型交通調査の結果、拡大推計された OD 表について比較を行った。おおむね高い相関が得られた。ただ、課題もあり OD 算定の基礎となるトリップは、出発地から目的地のペアを数えるものであるが、CDR では移動がずっと切れ目なく補足されることから移動としては正確であっても、どこで移動を切断し、トリップとするかという点で不確実性があり、その点での不一致が発見された。一方、災害対応(洪水対応)については雨期にほぼ水没することで有名な場時プール地域では雨期・乾期の人口差が著しいことがわかるなど、いわゆる避難誘導だけでなく、人々のリスク回避的な生活パターンまでが読み取れることがわかった。しかしながら、2016年7月のダッカにおけるテロ事件の結果、上記交通調査チームの多くの方が犠牲となり、またその後渡航が不可能になったことにより、これ以上の深い分析を行うことができなかつた。亡くなられた方の冥福を祈ります。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 2件)

Arai, A., Fan, Z., Matekenya, D., & Shibasaki, R. (2016). Comparative Perspective of Human Behavior Patterns to Uncover Ownership Bias among Mobile Phone Users. ISPRS International Journal of Geo-Information, 5 (6), 85. doi:10.3390/ijgi5060085

関本義秀, 櫻山武浩, 長谷川瑠子, 金杉洋, スパースな携帯電話通話履歴を用いたリンク交通量の推定~ダッカの事例 交通工学論文集,1,4,p1-8,2015/04/01

[学会発表](計 10件)

Fan, Z., Arai, A., Song, X., Witayangkurn, A., Kanasugi, H., & Shibasaki, R. "A Collaborative Filtering Approach to Citywide Human Mobility Completion from Sparse Call Records", Proc. of 25th International Joint Conference on Artificial Intelligence (IJCAI), pp. 2500-2506, 2016. <http://dl.acm.org/citation.cfm?id=3060971>

Dunstan Matekenya, Masaki Ito, Yoshito Tobe, Ryosuke Shibasaki, Kaoru Sezaki, "Enhancing Location Prediction with Big Data: Evidence from Dhaka," 4th International Workshop on Human Activity Sensing Corpus and Application co-located with ACM Ubicomp 2016, Sept. 2016.

[産業財産権]

出願状況(計 0件)

取得状況(計 0件)

6. 研究組織

(1)研究代表者

柴崎 亮介 (Ryosuke Shibasaki)

東京大学・空間情報科学研究センター・教授
研究者番号:70206126

(2)研究分担者

関本 義秀 (Sekimoto Yoshihide)

東京大学・生産技術研究所・准教授
研究者番号:60356087

瀬崎 薫 (Sezaki Kaoru)

東京大学・空間情報科学研究センター・教授
研究者番号:10216541

渡辺 知保 (Watanabe Chiho)

東京大学・大学院医学系研究科・教授
研究者番号:70220902

(3)連携研究者

金杉 洋 (Kanasugi Hiroshi)

東京大学・地球観測データ統融合連携研究機構・特任研究員
研究者番号:00526907

新井亜弓 (Arai Ayumi)

東京大学・空間情報科学研究センター・特任研究員
研究者番号:10788574

Witayangkurn Apichon (ウィタヤンクーン・アピション)

東京大学・空間情報科学研究センター・特任助教

研究者番号:90726407

(4)研究協力者

なし ()