

## 科学研究費助成事業 研究成果報告書

令和 6 年 6 月 7 日現在

機関番号：11301

研究種目：若手研究

研究期間：2021～2023

課題番号：21K14365

研究課題名（和文）犯罪がいつ起こるかを予測する：時期的犯罪予測の手法構築

研究課題名（英文）Predicting When Crime Will Occur: Building a Methodology for Timely Crime Prediction

研究代表者

大山 智也 (Ohyama, Tomoya)

東北大学・データ駆動科学・AI教育研究センター・助教

研究者番号：80893776

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：本研究では、既存研究で疎かにされてきた、犯罪発生時期を詳細に予測する手法の開発に取り組んだ。具体的には、先んじて把握可能な時間要因（特別日、曜日、天候、催事など）からの効果、犯罪企図者の活動が活発化することによる変動の2つを統合し、発生日の予測を試みた。大阪市（車上狙い・部品狙い）、ソウル市（屋外での暴力、性犯罪）で提案手法の予測性能を評価した結果、ランダムに要警戒日を選ぶモデルや、既存の機械学習手法を用いたモデルを凌駕した。また、単純な暴力よりも性犯罪での予測性能が高かったことから、提案手法は、明確な犯行意図をもって行われる犯罪においてより有効である可能性が示唆された。

研究成果の学術的意義や社会的意義

犯罪予測は、理論、データ、アプローチ含め、これまで地理的な側面（どこで起こるか）に力点がおかれ、時間的な側面（いつ起こるか）に関しては相対的に関心が低く、予測性能は限定的であった。そのため、犯罪が起こる地域・場所はわかっていても、時期がずれて防げないという事態が生じていた。本研究は、こうした問題を解決するために犯罪発生の時間的要素を詳細化し、曜日等の予め把握可能な要因と、犯罪時系列データさえあれば、ある程度の精度で予測が可能であることを示した点で意義がある。また、低頻度な犯罪は発生時期まで予測するのが困難であるが、本研究では予め地域を地理的にグルーピングして予測するアプローチの有効性も示した。

研究成果の概要（英文）：In this study, I developed a method to accurately predict the timing of crimes, an aspect that has been neglected in existing research. Specifically, I aimed to predict the date of occurrence by integrating two factors: (1) the effects of time factors that can be identified in advance (special days, days of the week, weather, events, etc.), and (2) fluctuations due to the increased activity of motivated offenders. I evaluated the prediction performance of the proposed method in Osaka City (vehicle theft and parts theft) and Seoul City (outdoor violence and sex crimes). The results showed that the proposed method outperformed models that randomly select days and models using existing machine learning methods. Furthermore, the prediction performance was higher for sexual crimes than for simple violence, suggesting that the proposed method may be more effective for crimes committed with clear criminal intent.

研究分野：犯罪学、社会工学

キーワード：安全・安心 犯罪予測 犯罪予防 時空間モデリング 地理的クラスタリング 異常検知 機械学習

### 1. 研究開始当初の背景

犯罪予測 (crime prediction) は、ひったくりや車上狙いのように、不特定多数の人・財物を対象とした犯罪について、翌月～翌日など比較的近い将来の発生リスクを、数百メートル長のグリッドセルや町丁目等の単位で予測し、地図上に示す技術である。

犯罪予測を初めて定義した Perry et al. (2013) は、これを「犯罪リスクの増大する場所・時を予測するアプローチ」とした。しかしながら、従来の犯罪予測研究は「いつ・どこで」のうち、「どこで」犯罪が起きるかに焦点を当てたものであり、「いつ」起きるかという視点は軽視されてきた。その結果として、予測パトロールの現場では、予測された場所は合っていた（高リスク領域として予測された実績があった）が、予測された時期がずれていたために犯罪を防げなかった (Ratcliffe et al., 2019) という事態がしばしば起こっていた。

これを回避するためには、空間ごとのリスクだけでなく、時間ごとのリスクを精緻に予測する必要がある。ただし、犯罪は低頻度事象であり、正例（たとえば犯罪が1件以上発生したユニット）が著しく少ない不均衡データになりがちである。そのため、機械学習にたえるデータとするために、数年～十年単位でデータを統合するといったことが行われてきた。これにはその期間、犯罪の発生傾向が変わらないという強い仮定をおく必要があった。

### 2. 研究の目的

本研究では、次の3つの課題を設定した。

#### 課題1 基礎的検討：季節・週内・日付・時間帯固有の犯罪発生リスクの推定

犯罪発生に対する、季節による効果、週の中の曜日（月、火…土、日）による効果、祝日や大規模イベント等開催日の日付による効果といった時期に固有の犯罪誘引効果を検討する。

#### 課題2 基礎的検討：ひとたび被害が起きてから上昇する犯罪発生リスクの推定

曜日等に固有の効果とは独立に、未検挙の犯罪企図者・集団が活動し続けることで上昇する犯罪反復効果を検討するため、ひとたび犯罪被害が発生した直後の数日～数週間以内に急上昇する被害水準に着目し、期待される持続期間がどの程度なのかを推定する。

#### 課題3 基礎的検討の統合による予測手法の具体化：時期的犯罪予測の手法構築と精度検証

課題1, 2の成果から、日付や時間帯に固有の犯罪誘引効果と、被害発生後急上昇する犯罪反復効果とを統合し、犯罪の発生する日・時間帯を町丁目ごとに予測する手法を構築、過去のデータを用いながら実際の犯罪発生状況と照らし合わせ、予測精度を検証する。

課題1～3にあたって、不均衡データへの対処として、地域の諸属性に着目したグルーピングを行い、犯罪予測の時間的詳細化を試みた。本研究は、地理的な側面に偏っていた犯罪予測研究を時間的側面に拡張するものである。また、犯罪が相対的に低頻度な日本において時間的に高精度な予測ができれば、海外においても凶悪犯罪などの低頻度罪種の予測に援用でき、犯罪予測研究にも示唆を与えることが期待できる。

### 3. 研究の方法

ここからは、実際の予測手法を日本（大阪市）と韓国（ソウル市）の2つに適用した分析について記す。

#### (1) 大阪市での分析

まず、大阪市での分析について述べる。大阪市の人口は2018年時点で2,725,006人、面積は225.32km<sup>2</sup>であった。分析の基礎的単位は町丁目であり、1,904あった。その面積の中央値は0.091km<sup>2</sup>であった。

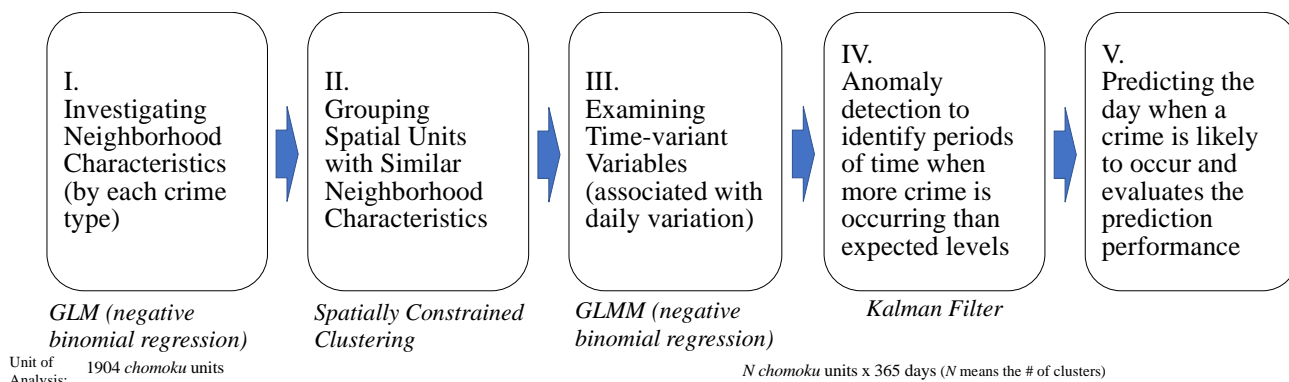


Fig. 1 The structure of current analysis

本研究の分析は、Fig. 1 に示す手順で構成される。まず (Step I)、社会経済的、物理構造的変数から車上狙い・部品狙いに関連する要因を抽出するために回帰分析を行った。応答変数となる刑法犯認知件数は負の値を除いた整数から構成されるカウントデータであり、かつ歪度の高い分布を示すことから、負の二項分布を仮定したモデリングを行った。ここでは全ての町丁目分析単位として回帰分析を実行、 $\alpha=0.05$  水準で統計的に有意なパラメータを同定した。

分析対象とした犯罪データは、2018、2019 年に発生した車上狙い、および部品狙いを統合した罪種 (以下、車上狙い等) とした。分析対象となった認知件数は 2018 年で 5,079 件、2019 年で 3,688 件であった。このうち、2018 年のデータを Fig. 1 における I~III における応答変数として、2019 年のデータを IV と V で用いた。

次に (Step II)、クラスター分析により町丁目グループをグルーピングした。上述の回帰分析により抽出した各種の犯罪に関連する変数をもとにクラスター分析を実行した。これにより、町丁目グループは各種の犯罪に関連する社会経済的、または物理構造的な特徴が似通った地理的集合に分けられる。ただし、通常のクラスター分析を行った場合、町丁目グループ間の空間的な隣接関係を無視して離れた町丁目グループを同じクラスターに分類することになり、都市をモザイク状に分割してしまうことになる。そのため、ここでは隣接関係を考慮したクラスター分析を行った。これには、Esri 社が提供する標準的な GIS ソフトウェア ArcGIS PRO (ver 2.9.1) の“spatially constrained multivariate cluster analysis”を用いた。クラスターの最大サイズは  $20 \text{ km}^2$  とした。

その後 (Step III)、クラスター分析で見出された空間的なグループごとに、日~土の曜日、特別日、天候といった時変のダミー変数で日毎の犯罪発生件数を回帰する分析を行った。これには、2019 年の各日の犯罪発生件数を予測する目的だけでなく、犯罪発生と関連する予め把握可能な時変要因を同定する目的もある。分析では、町丁目数 $\times$ 365 日を分析単位として、応答変数である犯罪の認知件数について、誤差項が負の二項分布にしたがうことを仮定した回帰分析を行った。ただし、同一の町丁目内の一日ごとのデータは、必ずしも互いに独立とはいえない。この場合、independent and identically distributed (i.i.d.) の仮定にしたがっておらず、パラメータの有意性を検定する際に Type I error を犯す可能性が高まる。こうした事態に対処するため、本分析では個々の町丁目の切片に変量効果を仮定した generalized linear mixed model (GLMM) を適用し、階層性を考慮した分析を行った。

Step IV では、カルマンフィルタを利用した異常検知を行った。本分析では、2019 年の観測値を用いて推定された状態 (トレンド) に対し、Step III で得た時変変数による予測値から推定される状態を比較し、後者が前者の 2 倍以上の水準に達したタイミングを異常な時点とした。こうして検知されたタイミングは、曜日や天候といった予めわかっている要因から予測される水準から逸脱したものであり、犯行企図者のような不確定要素がはたらくことによる状態の変化を示すと考えることができる。カルマンフィルタのための参照期間は過去 2 週間として、2019 年の 1 月 1 日から 12 月 31 日までの 1 日ずつ、異常の判定を行った。

最後に (Step V)、Step II で分けられた町丁目グループごとに、2019 年の犯罪発生日を予測し、実際の発生状況との比較を行うことで、予測手法の性能評価を行った。予測にあたっては、まず、Step III で得られた時変係数による予測値の高い順に、上位 10% (36 日間) を警戒すべき日としてフラグを立てた (予測 v1)。次に、Step IV で異常と判定された日にフラグを立てた (予測 v2)。この v2 では、フラグが立った日数は、異常が検知されたタイミングの多さにより変動するため、クラスターごとに異なる。最後に、ランダムに 36 日を選んでフラグを立てる方法 (v0) で予測を行い、評価のベンチマークとした。評価指標は recall:  $TP/(TP+FN)$ , precision:  $TP/(TP+FP)$ , その調和平均である F1 スコアとした。

## (2) ソウル市での分析

ここからは、韓国 (ソウル市) での分析について述べる。人口は 2023 年時点で約 940 万人、面積は約  $605 \text{ km}^2$  である。分析単位は Dong (洞) と呼ばれる行政単位であり、これは国勢調査の集計値が算出される基礎単位でもある。今回対象としたソウル市内の 425 洞であり、その面積の中央値は  $0.974 \text{ km}^2$  であった。

分析対象とした犯罪データは、2018 年、および 2019 年に発生した単純暴行 (simple assault: 2018 年は 49,741 件、2019 年は 45,682 件) と性加害 (sexual offense: 2018 年は 4,357 件、2019 年は 4,317 件) であった。単純暴行は、事情と関係なく暴力をふるって怪我などの被害がある場合、このカテゴリに含まれる。ただし、強盗や殺人・殺人未遂などの凶悪犯罪や家庭内暴力、恐喝、脅迫などは含まれない。性加害は、セクハラ、性暴力など性を媒介として相手の意思に反してなされる全ての加害行為が該当し、公然わいせつなども含まれる。データは韓国の Police Science Institute より提供された犯罪認知件数データであった。なお、分析においては、犯罪予測とパトロールによる抑止可能性に鑑み、路上で発生した事象に限定した。

以降の分析の流れは、大阪市の分析と共通している。ただし、ソウル市の場合、Fig 1 における分析単位: Unit of Analysis の数は 425、クラスターの数は単純暴行で 42、性加害で 38 となった

(Step IIにおけるクラスターの制限サイズは 50 km<sup>2</sup>とした)。また、最終モデル(Step V)の評価方法については、予測にあたっては、まず、Step Iの回帰分析において得られる、社会経済的/物理構造的変数に基づいた犯罪発生件数予測値をもとに、その値の高い 127 洞(425 の Dong の上位約 30%にあたる)を抽出した。この 127 洞について、次の 2通り(予測 V1, V2)に分けて犯罪発生が予測される日を選んだ。予測 V1 では、Step III で得られた時変変数に基づいた日ごとの犯罪発生件数予測値をもとに、365 日×42 洞(上位 10%)を選んだ。予測 V2 では、Step IV で異常として検知された日を毎日 42 洞に達するまで選び、42 に達しない日は、残りの Dong を Step III の時変変数による予測値の高いものから選んだ。なお、ベンチマークとして、社会経済的/物理構造的変数と時変変数を区別せず、425 洞×365 日の 2018 年データ全てを一貫通貫にランダムフォレストで学習させ、2019 年の犯罪発生日を 42 洞×365 日で予測したもの、365 日間毎日ランダムに 42 洞を選ぶランダムセレクションの 2つを用意し、比較対象とした。

本研究における分析は、Step II 以外は統計解析環境 R ver 4.0.2 を使用し、Step I の回帰分析は MASS パッケージの glm.nb 関数、Step III の時変変数による回帰分析は lme4 パッケージの glmer.nb 関数を、Step IV の異常検知には KFAS パッケージの SSMModel, fitSSM, KFS 関数をそれぞれ用いた。

#### 4. 研究成果

大阪市、ソウル市の分析とともに、Step I から IV までの結果は紙幅の関係から割愛する。大阪市の分析結果として、Table 1 に最終的な予測の評価結果を示している。時変変数による予測と、異常検知による予測を組み合わせた v2 でも、recall ではせいぜい 0.5 程度、F1 スコアは 0.3 未満という結果となり、良好なパフォーマンスとはいえないものであった。また、v2 の F1 が上位 3 位までのクラスター 7, 11, 26 は、recall のわりに precision が高く、保守的な予測(たとえばクラスター 7 なら、予測フラグが立った日に警戒をすれば 9 割方、車上狙い等が発生するが、それ以外の日で 8 割以上の犯罪を見逃すことになる)となっていることがうかがえる。

予測のバージョンで、v1 と v2 の違いを見ると、ほとんどの場合で F1 は後者の方が高く、recall や precision も同様である。ランダムな予測と比べても、v1 は v2 よりもやや劣る結果となり、Step IV の異常検知による予測を加えた方が、予測パフォーマンスが高くなると考えられる。大阪市では、時変かつ静的な曜日や休日、季節、気候、催事といった要因のみから犯罪発生時期を予測するのが困難であったといえる。ただし、それ自体が無用であるわけではなく、そうした要因から予測される水準を犯罪発生の実態が上回っているときに警戒を行うことで抑止が可能であることも、今回の結果は同時に示唆する。

Table 2 には、ソウル市の分析の結果となる提案モデル V1, V2 による予測と、ベンチマークとなる単純なランダムフォレスト(RF)、ランダムセレクション(RS)による予測を比較した結果を示した。概して、提案手法の recall は、単純なランダムフォレスト回帰(RF)を上回った。特に、比較的発生頻度の低い性加害については、提案モデル V2 は RF を 0.1 以上の水準で上回り、recall は RF の 0.286 に対して 0.401 であった。これは、提案モデルが RF よりも 12%近く多くの犯罪発生日を予測できることを示しており、この差は警察の実務に大きな影響を与えよう。単純暴行では、この差は小さい。また、性加害では、異常検知を用いた V2 では V1 に比べて約 0.02 改善するが、単純暴行ではその差はほとんど無視できる。これは、性加害のメカニズムに関連している可能性がある。酔客の喧嘩など含まれる単純暴行とは対照的に、性加害では、明確な犯意を持った犯罪者がターゲットを探し、逮捕されずに活動を続けた場合に、活動頻度がエスカレートするなどして、傾向が変動する可能性がある。今回の結果は、V2 が予測にこのような傾向をうまく利用できた可能性を示唆している。なお、性加害については precision が低い。これは不均衡データの場合にはままあることで、RF と RS でも共通している。

考察として、まず、大阪市の分析においては、提案手法はランダムな予測を概ね上回るパフォーマンスをあげていたものの、クラスターによっては差が小さいものもあった。対してソウル市では、予測性能の評価方法は少し異なるものの、提案手法はランダム予測や、単純な機械学習モデルに比して良好なパフォーマンスであり、その差は単純暴行よりも、頻度の低い性加害でむしろ大きかった。犯罪発生頻度については、大阪市の車上狙い・部品狙いとソウル市の性加害は同程度の水準(3,000~5,000 程度)であることから、提案手法の性能が大阪市(日本)で相対的に低かったのは、頻度よりも犯罪の種別による差によるものかもしれない。もっとも、地域性、文化・社会的背景といった犯罪を取り巻く環境の違いによる可能性もあり、今後、似通った罪種での検証が望まれる。

また、大阪市、ソウル市の両者において、予め把握可能な時変変数のみ用いた V1 モデルよりも、そこから予測される水準を大きく外れた変動を捉え、それを予測に取り入れた V2 モデルが例外なく高いパフォーマンスを発揮していた。このことは、犯罪発生水準が直近で犯罪が発生した場合に一時的に上昇することを主張する近接反復被害等(Townsley et al., 2003 など)の既存の犯罪学理論とも一貫するものである。ただし、本研究で用いた異常検知の手法は、よ

り一般的に犯罪者の活動水準の上昇を捉えたものであり、今回の結果は、こうしたトレンドを予測に取り入れるための一つの有効な方法を提案できたといえよう。また、分析を通じて、クラスターごとの時変変数が犯罪と関連しているかが異なっていることが示された。こういった傾向について考察できることは、犯罪抑止に大きな意味を持つだろう。

最後に、今回は町丁目をまとめたクラスターごとに犯罪発生日の予測が行われたが、クラスターによって広さ・含まれる町丁目数は異なり、実用にあたっては、警察署や交番の配置も考慮しながら警戒すべき日・場所を定めてゆく必要がある。この点は、実装に関連した課題として今後新たに研究を進めてゆく必要があろう。

Table 1 大阪市における予測性能評価結果

	recall_v0	recall_v1	recall_v2
Min.	0.000	0.000	0.000
Median	0.080	0.091	0.170
Mean	0.084	0.101	0.210
Max.	0.333	0.333	0.500
NA's	2	2	2
	precision_v0	precision_v1	precision_v2
Min.	0.000	0.000	0.000
Median	0.151	0.125	0.150
Mean	0.213	0.221	0.235
Max.	0.811	0.923	0.911
NA's	0	0	0
	F1_v0	F1_v1	F1_v2
Min.	0.036	0.036	0.022
Median	0.131	0.129	0.148
Mean	0.122	0.126	0.150
Max.	0.226	0.261	0.281
NA's	10	8	4

Table 2 ソウル市における予測性能評価結果

	Model Performance (mean for 365 days)	Simple assault	Sexual offense
Recall	V1: by Time-variant Variables	0.324	0.379
	V2: V1 & Anomaly Detection	0.332	0.401
	Random Forest	0.309	0.286
	Random Selection	0.177	0.197
Precision	V1: by Time-variant Variables	0.477	0.084
	V2: V1 & Anomaly Detection	0.465	0.077
	Random Forest	0.452	0.064
	Random Selection	0.298	0.046
F1	V1: by Time-variant Variables	0.383	0.139
	V2: V1 & Anomaly Detection	0.384	0.130
	Random Forest	0.364	0.110
	Random Selection	0.220	0.086

<引用文献>

Corcoran J., Wilson I., and Ware J. (2003) Predicting the geo-temporal variations of crime and disorder, *International Journal of Forecasting*, **19**(4), 623-634.

Farrell, G., and Pease, K. (1993) *Once bitten, twice bitten: Repeat victimisation and its implications for crime prevention.*, London: Home Office.

Ivaha C., Al-Madfai H., Higgs G., and Ware J. (2007) The Dynamic Spatial Disaggregation Approach: A Spatio-Temporal Modelling of Crime, *World Congress on Engineering*, 961-966.

Johnson S., Bowers K., Birks D., and Pease K. (2009) Predictive mapping of crime by ProMap: accuracy, units of analysis, and the environmental backcloth, in *Putting crime in its place: units of analysis in spatial crime research*, D. Weisburd, W. Bernasco, and G. Bruinsma, Eds. New York: Springer, pp. 171-198.

Mohler G., Short M., Brantingham P., Schoenberg F., and Tita G. (2011) Self-exciting point process modeling of crime, *Journal of the American Statistical Association*, vol. **106**, 100-108.

Perry W., McInnis B., Price C., Smith S., and Hollywood J. (2013) *Predictive policing: The role of crime forecasting in law enforcement operations*, Rockville: NCJRS Photocopy Services, xiv pp.

Ratcliffe J., Taylor R., and Fisher R. (2019) Conflicts and congruencies between predictive policing and the patrol officer's craft, *Policing and Society*, **30**(6), 639-655.

Townsley, M., Homel, R., and Chaseling, J. (2003) Infectious burglaries. A test of the near repeat hypothesis. *British Journal of Criminology*, **43**(3), 615-633.

Zhu Q., Zhang F., Liu S., and Li Y. (2019) An anticrime information support system design: Application of K-means-VMD-BiGRU in the city of Chicago, *Information & Management*, **103247**.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 3件）

1. 著者名 Tomoya Ohyama	4. 巻 7
2. 論文標題 Estimating Time-Variant Variables' Effects on a Micro Scale for Time-Based Crime Prediction	5. 発行年 2022年
3. 雑誌名 Proceedings of 7th International Conference on Business and Industrial Research	6. 最初と最後の頁 143-148
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Tomoya Ohyama, Kazunori Hanyu, Masayuki Tani, Momoka Nakae	4. 巻 130
2. 論文標題 Investigating crime harm index in the low and downward crime contexts: A spatio-temporal analysis of the Japanese Crime Harm Index	5. 発行年 2022年
3. 雑誌名 Cities	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.cities.2022.103922	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Tomoya Ohyama, Ilgon Kim, and Kwang-Ho Jang	4. 巻 9
2. 論文標題 A Time-Focused Approach to Crime Prediction: An Analysis in Seoul.	5. 発行年 2024年
3. 雑誌名 Proceedings of 9th International Conference on Business and Industrial Research	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件／うち国際学会 2件）

1. 発表者名 大山智也・羽生和紀・谷真如・中江百花
2. 発表標題 市民の認知する犯罪危害の地理的分析：47都道府県パネルデータを用いた検討
3. 学会等名 第31回地理情報システム学会研究発表大会
4. 発表年 2022年

1. 発表者名 Tomoya Ohyama
2. 発表標題 Estimating Time-Variant Variables' Effects on a Micro Scale for Time-Based Crime Prediction
3. 学会等名 2022 7th International Conference on Business and Industrial Research (ICBIR) (国際学会)
4. 発表年 2022年

1. 発表者名 大山智也・金一坤
2. 発表標題 犯罪発生時期はどこまで予測できるか 時空間的犯罪予測の試み
3. 学会等名 第32回地理情報システム学会研究発表大会
4. 発表年 2023年

1. 発表者名 Tomoya Ohyama and Ilgon Kim
2. 発表標題 When Does Crime Occur? : A Time-Focused Approach to Crime Prediction.
3. 学会等名 American Society of Criminology Annual meeting 2023 (国際学会)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	金 一坤  (Kim Ilgon)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
韓国	Police Science Institute of Korea			