

Data-derived Green's function method による犯罪予測

2017/01/17 VLED 第2回利活用・普及委員会

梶田真実

(collaboration with 梶田晴司)

アジェンダ

1. 自己紹介と背景

2. 犯罪予測

2-1. 背景と着眼

2-2. 独自手法「DDGFアルゴリズム」

2-3. シカゴ, 東京のオープンデータを用いた デモンストレーション

3. 展望

1. 自己紹介と背景

自己紹介

梶田 真実 (主婦)

略歴：

千葉大学飛び入学→2010年3月東京大学大学院 Ph.D

→学術振興研究員(PD)で大阪大学、名古屋大学

→イタリア→トビラシシステムズ→東京大学空間情報科学研究センター

犯罪予測アルゴリズムの研究をしています。

専門分野：

- ・ 統計物理学、理論物理学、非線形動力学
- ・ ガラス転移などフラストレーションのある系の物理

手法：

- ・ 特異摂動法を用いた確率過程ダイナミクスの粗視化、
時空間特徴量の解析
- ・ 統計物理学的モデリング

主な業績

複雑なフラストレーションのある多体系の確率過程に対して、解析的に時空間特徴量を抽出する粗視化手法を独自に構築した。

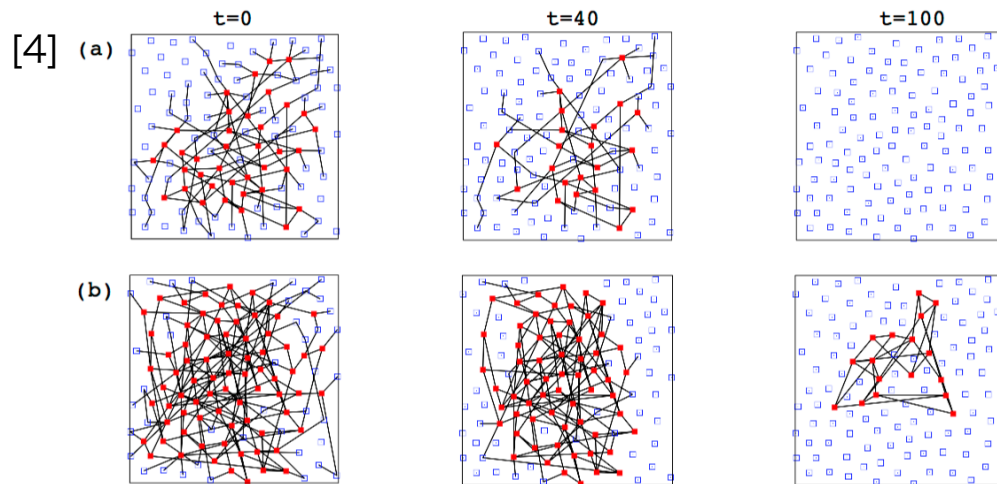
(背景) 渋滞、ガラス転移点近傍での協働現象などのカスケード現象には、臨界的な揺らぎが発生する。

(着眼) 特異摂動法を用いたダイナミクスの粗視化により、ある種のカスケード現象に本質的な秩序変数を数理構造から抽出することに成功

(手法)

カスケード現象のダイナミクスに対して、サドルノード分岐点近傍で揺らぎのある系という構造を特異摂動を適用し手法を開発[2][6]

幾つかのモデルに対して（密度場FP方程式[1][3]、kcoreパーコレーションモデルの動力学[4]、モード結合方程式[5]）この手法を適用し、秩序変数の導出、臨界指数の計算に成功。エイジング現象、降伏応力の発生の起源への足がかりに。



論文

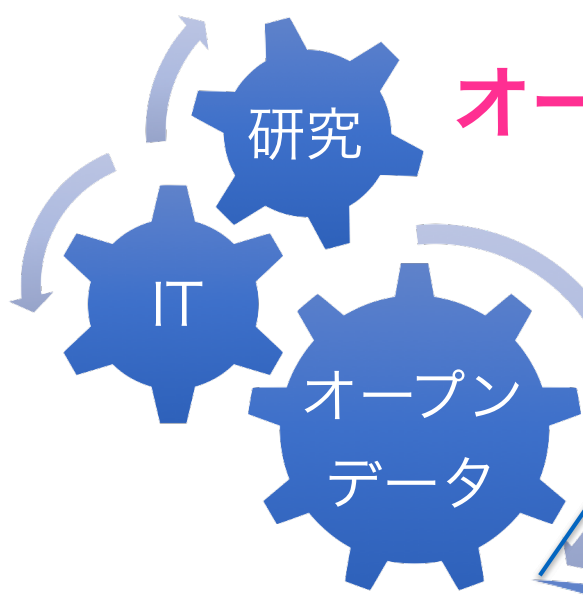
- [1] J. Stat. Mech. L10003 (2006)
- [2] Europhys. Lett. 77 50008 (2007)
- [3] Phys. Rev. E 78, 055202(R) (2008)
- [4] J. Phys. A: Math. Theor. 42, 075005 (2009)
- [5] J. Phys. A: Math. Theor. 42, 245001 (2009)
- [6] Phys. Rev. E 82, 011127 (2010)

競争的外部資金獲得

平成20年4月～平成22年3月 公益信託林女性自然科学者研究助成基金研究奨励「林フェロ一」助成金
平成22年4月～平成25年3月 科学研究費補助金(平成22年度学振特別研究員奨励費)

$$[5] \partial_t \phi(t) = -\phi(t) - g \int_0^t ds \phi^2(t-s) \partial_s \phi(s)$$

オープンデータ + アプリケーション開発の履歴



家計簿アプリ「Pockets」



「ガソリン天気予報」



原油先物時系列データ



「国会議事録アプリ」よむコッカイ

衆議院国会議事録、
参議院国会議事録、

wikipedia

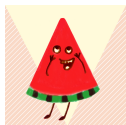


「パトロールマップ」

各都道府県警察配信ML
の軽犯罪情報

犯罪予測アルゴリズム

「Data-derived Green's function method」



地図にイコト、フルイコト
を見やすく表示



防犯アプリ「パトロールマップ」



週刊アスキー様による1分紹介：<https://www.youtube.com/watch?v=Gz6Nceqdy0I>

よりご紹介いただきました
Refs.

週刊アスキー様 熱血Appバトル【第4回審査会】 <https://www.youtube.com/watch?v=k1nO3yVGx9M>

Patrolmapの使い方：<https://www.youtube.com/watch?v=annwfTdxFxs>

次世代型の防犯は「事後対応」から「予測」へ

現状

- ・ 警察署からのメーリングリスト、アプリで過去に発生した犯罪情報は公開されている
- ・ 防犯ボランティア団体の地道な活動で近年犯罪認知件数の減少を記録した県もある

課題：「カン・コツ」パトロールのみでは人的コストが高い／公開犯罪データが限定的・非活用
→ より効果的な防犯に向けて、データ分析による有効活用の余地あり

世界では既に「予測」による犯罪抑止へとシフトし始めている

1. PredPol^[1] in USA



2011年にGeorge Mohler准教授が提案したEM法を用いた**犯罪予測アルゴリズム**[2]をベースに、警察向けパトロール経路提案事業のスタートアップ Predpolが生まれた。
これを用いたパトロール活動により**犯罪発生が20%程度減少した**という目覚ましい成果報告がある[3]。

LA,シアトル、サンタクルーズ等の警察が導入

2. PreCobs^[4] in Germany



- [1] <http://www.predpol.com>
- [2] Mohler, G.O., et al., 2011. J. American Statistical Association, 106 (493), 100-108
- [3] <http://cooeydailypress.com/can-police-foretell-crime/>
- [4] <http://www.ifmpt.de>

バイエルン州の行政機関が導入

目的

高精度の犯罪予測アルゴリズムを考案する。

→より効果的なパトロール経路を提案するなど、
防犯サービスのための基礎技術を構築する。

2. 犯罪予測

2-1. 背景と着眼

2-2. 独自手法「DDGFアルゴリズム」

2-3. シカゴ, 東京のオープンデータを用いた
デモンストレーション

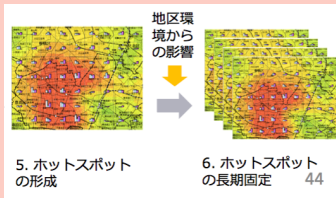
2-1. 背景と着眼



Short, Tita, Brantingham, Chayes et al
Math Mod Meth Appl Sci(2008).

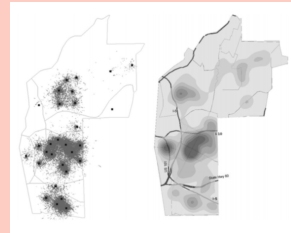
地理統計分析
空間情報科学

犯罪の地理的分析
道路のリンク接続関係を利用したネットワークの
点事象集積検出法



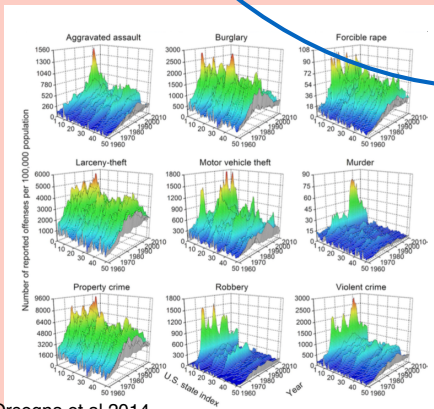
Amemiya et al. 2014資料より抜粋

Ryo Inoue 2013



犯罪のホットスポットの時空間分布の分析
環境要因の分析（罪種毎件数、施設、地理構造、
人口等の独立変数）
犯罪集中地区の移動パターン、環境要因の分析
犯罪発生マップの2次元カーネル密度推定

GIS



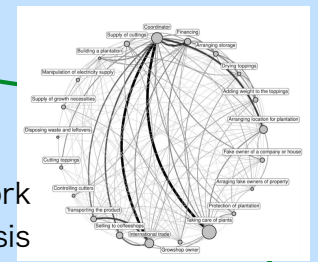
Orsogna et al 2014

犯罪種別毎の字空間分布の分析

プロファイリング

犯罪学
犯罪心理学

犯罪予測

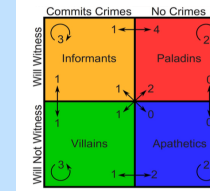


Network
Analysis

Hegemann, Tita et al Physica A (2011).

数理モデル

Social dilemma



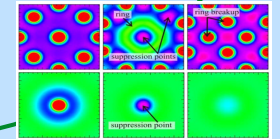
ポイントプロセスモデル

$$\lambda(t, x) = \sum_{t_i < t} g(t - t_i, x - x_i) + \lambda_0$$

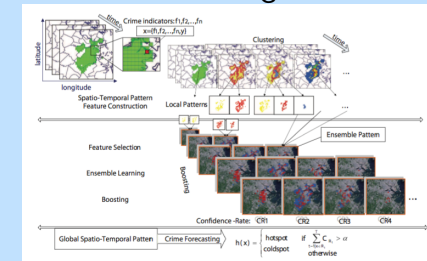
パターン形成 based models

$$\frac{\partial B}{\partial t} = \frac{\eta D}{z} \nabla^2 B - \omega B + \epsilon D \rho A,$$

$$D = \ell^2 / \delta t, \epsilon = \delta t, \text{ and } \rho(s, t) = n_s(t) / \ell^2$$



Ensemble Learning

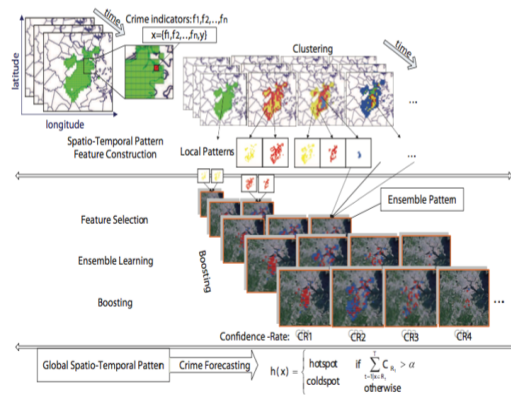


Chung-Hsien Yu¹, Wei Ding¹, Ping Chen¹, and Melissa Morabito 2010

機械学習

犯罪予測への数理モデルアプローチ

Ensemble Learning



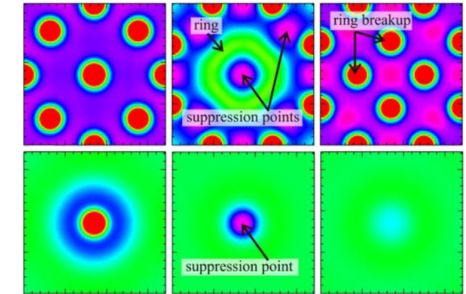
Chung-Hsien Yu¹, Wei Ding¹, Ping Chen¹, and Melissa Morabito 2010

パターン形成のモデル

$$\frac{\partial B}{\partial t} = \frac{\eta D}{z} \nabla^2 B - \omega B + \epsilon D \rho A,$$

$$D = \ell^2 / \delta t, \epsilon = \delta t, \text{ and } \rho(s, t) = n_s(t) / \ell^2$$

ex) パラメトリックな手法



Short et al. 2008

機械学習はデータ数必要

数理モデルだけだと実用化に遠い

データ駆動型+ノンパラメトリック型数理モデル



ポイントプロセスモデル

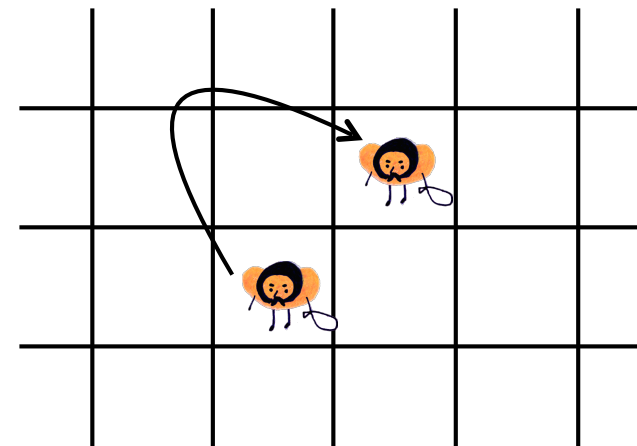
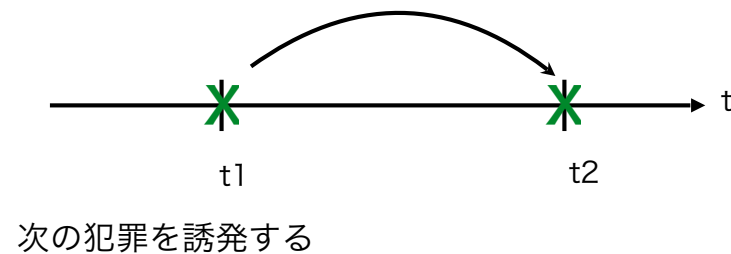
$$\lambda(t, x) = \sum_{t_i < t} g(t - t_i, x - x_i) + \lambda_0$$

近接反復被害

“犯罪者は犯行現場に帰ってくる”

環境とターゲットの行動を熟知した現場近くで犯行を繰り返す

犯罪のカスケード現象



犯罪には時空間相関がある

例：住居侵入が起きるとその2, 6日後、
1m圏内で再発生確率が高くなる


Ref: G.Mohler, M.Short et al 2011[2] Journal of the American Statistical Association,106, 100

犯罪予測の数理モデル ～ポイントプロセスモデル～

犯罪発生密度 $\lambda(t,x)$:

t, x における単位時間・面積あたりの 犯罪発生数

過去の事象から直接誘発されない寄与

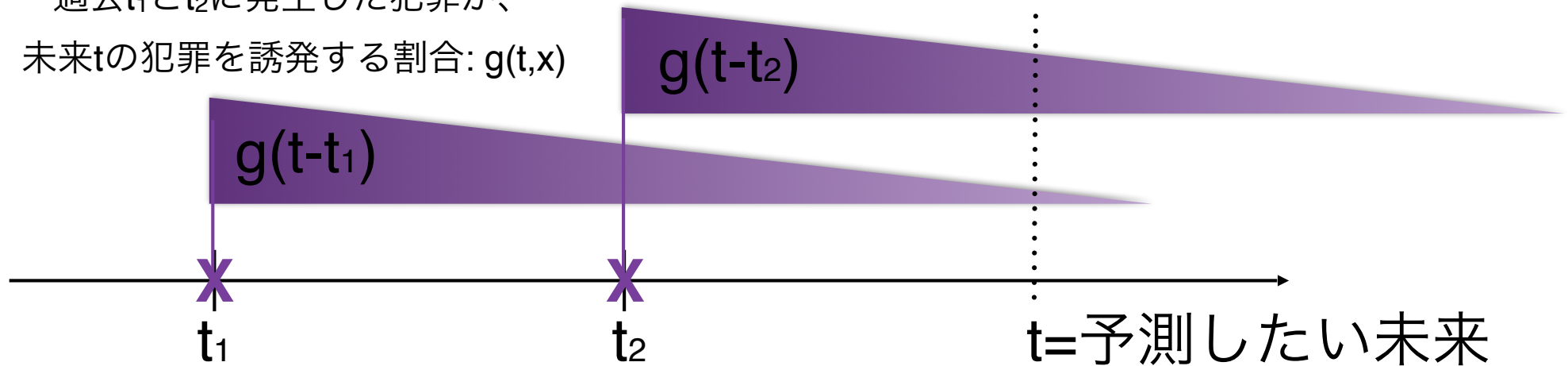
$$\lambda(t,x) = \sum_{t_i < t} g(t-t_i, x-x_i) + \lambda_0$$


過去の i 番目の犯罪イベントとの因果関係により誘発される寄与

[例]

過去 t_1 と t_2 に発生した犯罪が、
未来 t の犯罪を誘発する割合: $g(t,x)$

予測犯罪発生密度 $\lambda(t) = g(t-t_1) + g(t-t_2) + \lambda_0$



How to find $g(t, x)$?

How to find $g(t,x)$?

先行研究

prospective hot spot maps

$$g(t, x) = \frac{1}{(1+t)(1+x)}$$

と仮定

問題点：パラメトリック

Marsan method

データからExpectation Maximization algorithm(機械学習手法)を使って $g(t,x)$ を構築

[1] D.Marsan, O.Lenglin, Science 2008[1] Science, 319, 1076-1079.

[2] G.Mohler, M.Short et al 2011[2] Journal of the American Statistical Association,106, 100

問題点：精度に難あり
(特にデータが少ない場合)

目的

地域や罪種に依存しないノンパラメトリック
かつ精度の高い手法を開発する

2-1. Data-derived Green's function (DDGF)アルゴリズム

着眼点：近接反復を示す犯罪イベントは因果律を持つと考えられる。

→犯罪イベントを記述する何らかの偏微分方程式系がある

- ・ただし偏微分方程式の具体的な形は（もちろん）分からない
- ・分かるのは犯罪イベントのデータ、つまり解

→データから逆に偏微分方程式を決めることができるのでは

戦略：ポイントプロセスモデルを連続化。

**偏微分方程式の逆演算子である $g(t,x)$ (グリーン関数)
を用いて犯罪事象 ρ と関係付ける**

$$\rho(x, t) = \int_0^t dt' \int dx' g(t - t', x - x') \rho(x', t') + \int dx' g(t, x - x') \rho(x', 0)$$

犯罪データ ρ からグリーン関数 $g(t,x)$ を求めるという問題に帰着

アルゴリズムの概略

$$\rho(x, t) = \int_0^t dt' \int dx' g(t - t', x - x') \rho(x', t') + \int dx' g(t, x - x') \rho(x', 0)$$

データ生成型

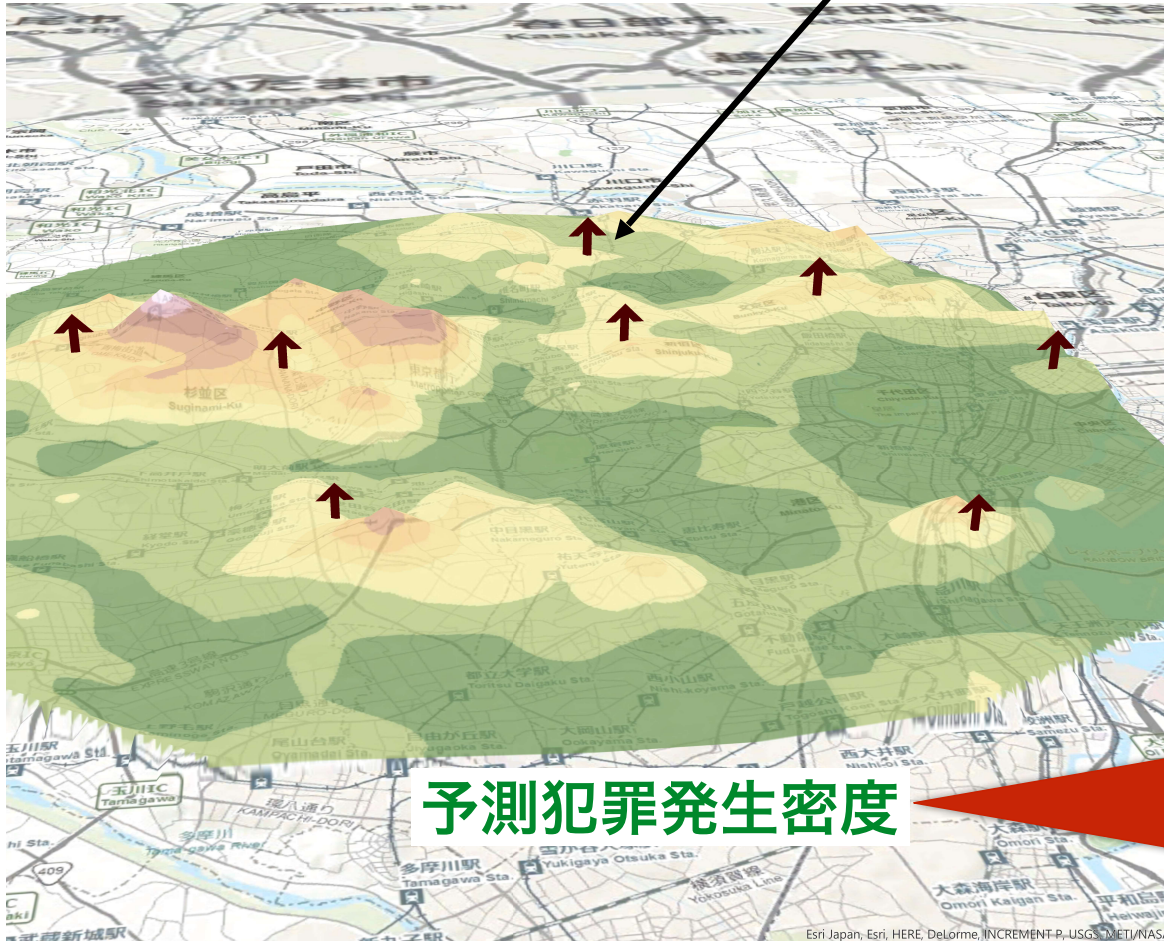
ノンパラメトリック型

- ① 方程式を解く。
- ② 過去のデータから $g(t, x)$ を構築。
- ③ $g(t, x)$ を使って、未来の犯罪を予測する

* 近日、学術論文にて公開予定

東京の予測犯罪発生密度と実際に起きた犯罪

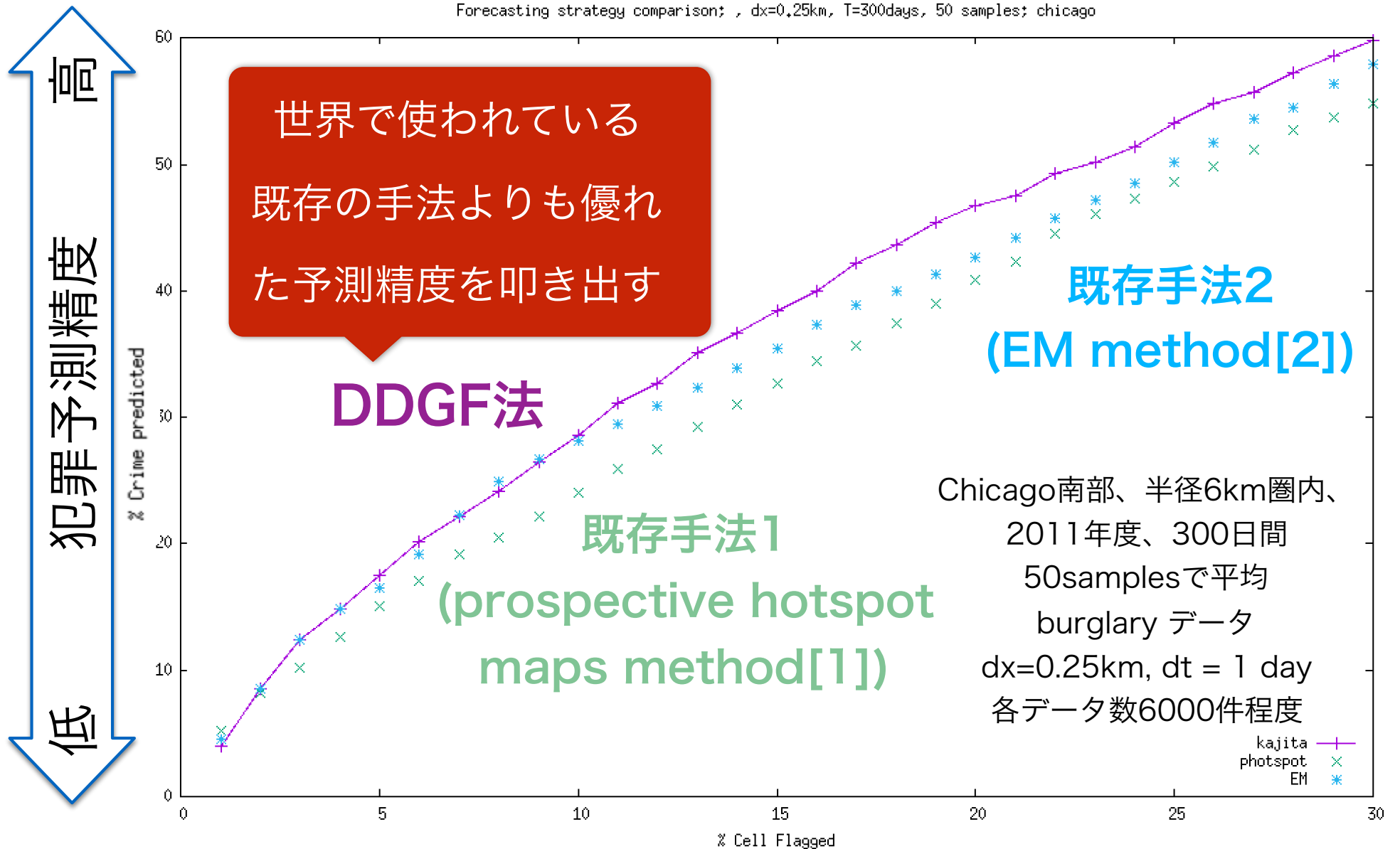
2015/7/2に実際に起きた犯罪



- ・ 警視庁のメーリングリストサービス（メール警視庁）が配信する東京都の軽犯罪データ
- ・ 東京都の痴漢、窃盗、不審者、暴漢
- ・ 東京都庁から半径10km圏内、2015/3月～2015/10月までの期間から100日間
- ・ 時空間分解能: 1km², 1 day

予測犯罪発生密度が高い
エリアで実際に犯罪が発

2-2. シカゴのオープンデータを用いた デモンストレーション



[1] Bowers, K.J., Johnson, S.D., and Pease, K., 2004. British Journal of Criminology, 44 (5), 641-658

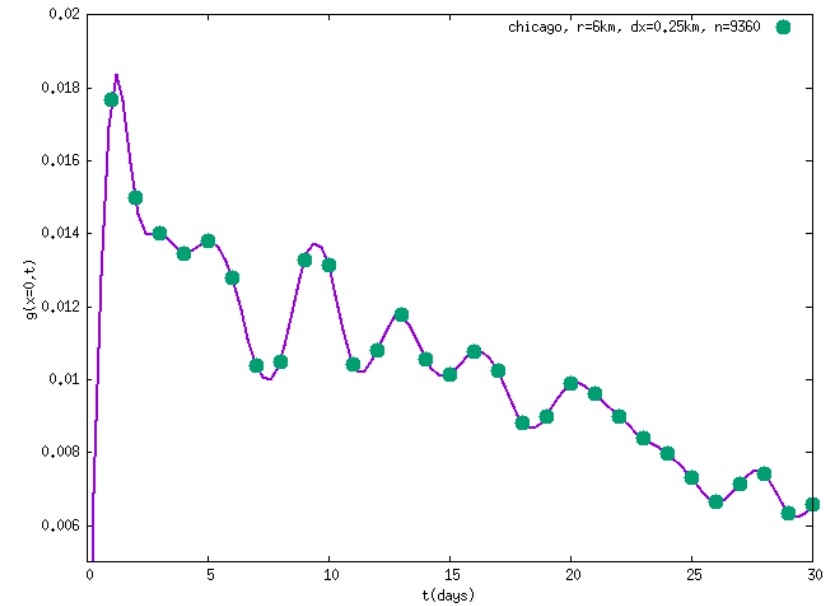
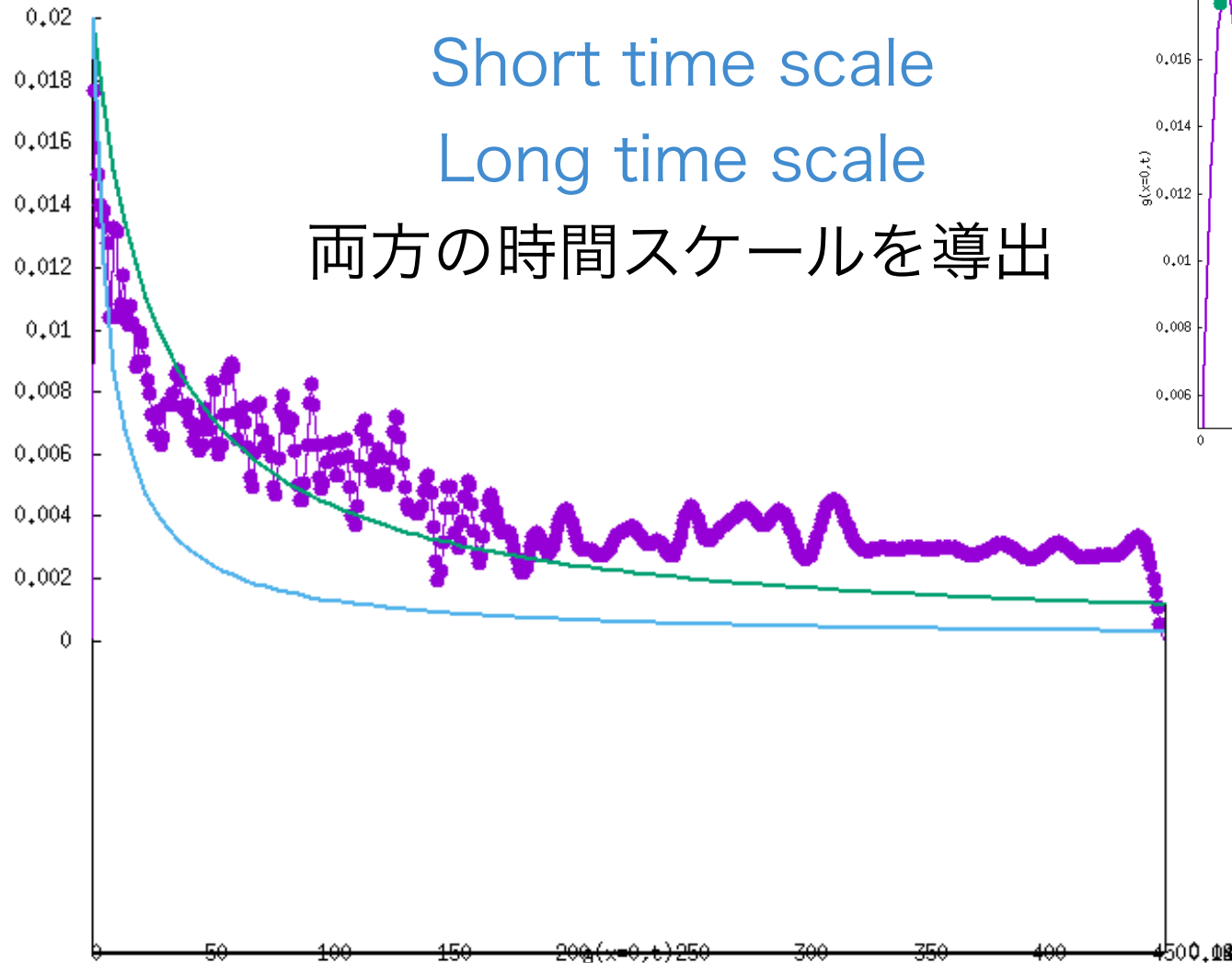
[2] Mohler, G.O., et al., 2011. Journal of the American Statistical Association, 106 (493), 100-108

DDGF法の特徴

Short time scale

Long time scale

両方の時間スケールを導出



Chicago南部、半径6km圏内、
2011年度、450日間, burglary
 $dx=0.25\text{km}$, $dt = 1 \text{ day}$
各データ数3200件程度

EM, PHOT法に比べて、long time tailが残るため
Causalな寄与とbackground的な寄与を自然に含む

3. 展望

(希望的) 展望

- ・ 警察と大学との共同研究 for 実証実験 in Japan.
→ 日本版PredPolサービス。

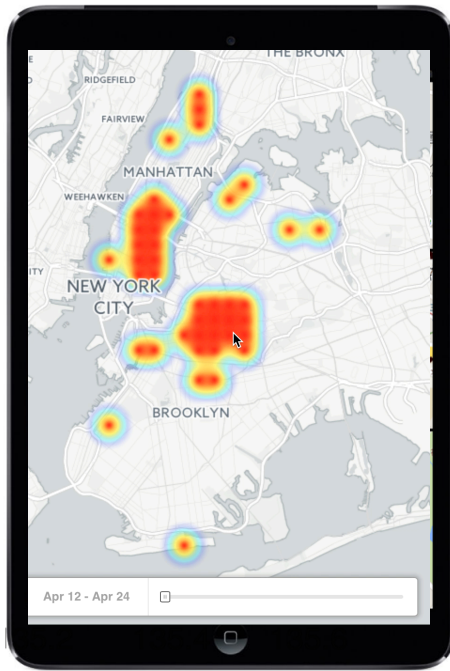
市民と警察とをつなぐ、情報シームレスの先端防犯プラットフォームを構築できる。

嬉しさ

効率良いパトロール経路を選ぶことで犯罪抑止に役立てられる。

アプリ上で

予測犯罪発生率を表示



犯罪発生情報

警察・自治体

DDGF法による
犯罪予測

パトロール経路提案

課題

- ・ 統計処理に十分な犯罪データが警察外部へ公開されていない。
- ・ 公開されているオープンデータは、非構造、多い欠損値、粗い分解能、低濃度データがほとんどで、分析に適さない。

展望（アイディア段階）

オープンデータを活用した「Research for Japan (R4J) 構想」

←(Ref.) Code for Japanの研究者バージョン

- ・ 企業に所属する研究者の中に、本業以外の研究で社会に貢献したい人がいる。
- ・ 細分化された分野／コミュニティで孤立した有志研究者を「for Japan」で繋げる）

嬉しさ

- ・ データのオープン化による技術・科学資産が創出
- ・ 社会実験がより広く・オープンに実施される。特定の民間企業に頼らずに課題解決。

オープンデータ とニーズ

嬉しさ

- ・ 業務外の研究に挑戦でき、論文も発表できる
- ・ 会社や組織の外からも評価される業績作り
- ・ 新しいキャリアチェンジへの足がかり

自治体・行政機関

サイクルが回るほど
日本が良くなる

R4J:オープン研究者コミュニティ

- ・ 研究者自体も組織に所属しながらも社会に貢献できる仕組（企業の副業解禁の流れを研究業界にも拡張）
- ・ 人材育成、雇用創出

ニーズを解決する 技術やサービス

- ・ 論文公開による社会の発展が促進される
- ・ シームレスなオープン研究コミュニティを形成
（理学、工学、社会科学の分野横断）
- ・ 新たな研究・事業領域が創出

Code4Japanのように、

参画するほど嬉しさが増幅する研究者のためのオープンプラットフォームが必要と考えている。