

サンプリングされた移動軌跡データと 感染シミュレーション

野田五十樹^{1,a)}

概要: サンプリングされた移動軌跡データを用いて感染シミュレーションを行う際に生じる課題と補正方法について検討する。ビッグデータと言えども、関係する人々の詳細な行動が全て記録されているわけではなく、何らかの形で部分的にサンプリングされたデータになっている。人をエージェントに対応付けるマルチエージェント社会シミュレーションでは、サンプリングされたデータを何らかの形で補う必要がある。よって、ビッグデータを用いたシミュレーションを行う場合、欠けている情報による影響を検討しておく必要がある。本稿では、感染シミュレーションにおいて移動軌跡のサンプリングデータを用いる際の、サンプリングの影響を実験及び数式により解析を試みる。また、その結果を用いて、欠けているデータを補う際の補正方法を検討する。

Statistical Feature of Sampled Trail Data and its Effect on Multiagent Pandemic Simulation

1. はじめに

シミュレーションは実世界のごく一部を切り出しているに過ぎない。計算機上に構築された仮想世界は、あくまで、シミュレーションで明らかにしたい現象を中心に実世界を切り出したものであり、かならず実世界と乖離がある。

シミュレーションで用いるデータについても同じである。ビッグデータと言えども、シミュレーションが対象としている世界の一部を切り取ったものに過ぎず、どうしても実際の現象とは乖離が生じる。そしてその乖離により、シミュレーションの結果も影響を受ける。

本稿では、ビッグデータと現実の差異がどのようにシミュレーションに影響を与えるかを、人の移動軌跡データを用いた感染シミュレーションを題材に分析を試みる。2019年から感染が始まったとされる COVID-19 は、現在、世界中で猛威をふるっており、それへの対処の手段の1つとして、各種の感染シミュレーションが試みられている [1], [2], [3]。中でも、人の移動と組合せた感染

シミュレーションは、経済活動シミュレーションとの連携の可能性もあることから、様々な形で取り組まれている。ただ、観測されたデータに基づくシミュレーションが、どの程度、現実を反映できており、また、生じてしまう違いについて補正をすることができるかを事前に知ることが重要である。

感染シミュレーションに活用できる人の移動データの1つとして、携帯端末の位置情報を記録したものがある。特に本稿では、携帯電話の ID 付きで時刻と位置データが記録された、移動軌跡データを取り上げる。このようなデータは、全人口あるいは全携帯端末をカバーしているわけではなく、ある一定の割合の人口をカバーしているサンプリングデータとなっている。また、個人情報保護の観点などから、位置情報も点としての位置データではなく、対象領域をメッシュ状のグリッドに分割し、どのグリッドに存在していたかで位置を表す形式を取ることが多い。つまり、人口カバー率に応じたサンプリングと、グリッド化による位置情報の粗視化により、実際の人々の行動からある程度の乖離が生じることになる。

以下では、このサンプリング及びグリッド化によるデータの乖離がどのように感染シミュレーションに影響を及ぼすかを分析し、シミュレーションにおける補正方

¹ 産業技術総合研究所 人工知能研究センター
AIST, Tsukuba, Ibaraki 305-8560, Japan

^{a)} i.noda@aist.go.jp

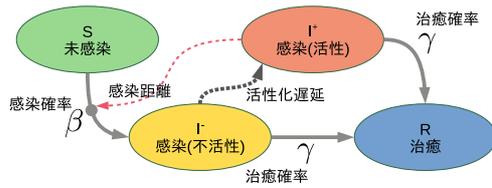


図 1 拡張 SIR 遷移モデル

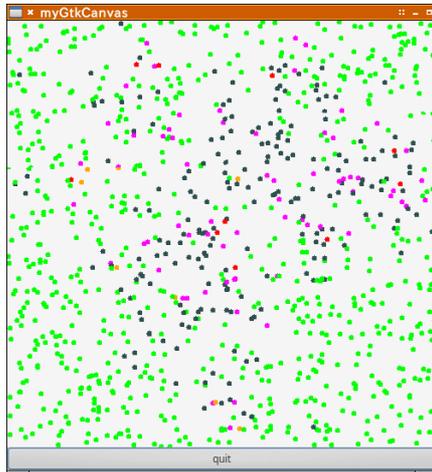


図 2 簡易感染シミュレーションの様子

法を導き出していく。本稿の構成は、まず 2 節で題材とする感染シミュレーションとサンプリングされたデータにより感染拡大が強調される影響を実験結果として示し、その原因を考察する。3 節では、その原因の一つのグリッド粗視化の影響の軽減方法を示し、実験により効果を確認する。つづいて 4 節では、もう一つの感染拡大の強調原因であるサンプリング復元について、人口分布の統計的性質に着目して数理的解析を行い、影響軽減のための補正方法を導出し、実験で確認する。5 節で提案手法と実験結果をまとめ、今後の課題などを考察する。

2. 簡易感染シミュレーション

2.1 シミュレーションモデル

分析のベースとする感染シミュレーションについて説明する。

本稿で用いるシミュレーションは、拡張 SIR 遷移モデルに人の移動モデルを組み合わせた、マルチエージェントシミュレーションである。個々のエージェントは人ひとりに対応する。拡張 SIR 遷移モデルとは、感染拡大のモデルである SIR モデルの状態遷移をベースに、感染状態を、不活性状態 (他者に感染を広げない状態) と活性状態 (感染を広げられる状態) に分け、不活性から活性への遷移は確率的ではなく時間経過に従って生じるものとする。すなわち、各エージェントの感染に関する状態遷移は以下のように生じる。

- 未感染状態: まだ一度も感染していない状態。この

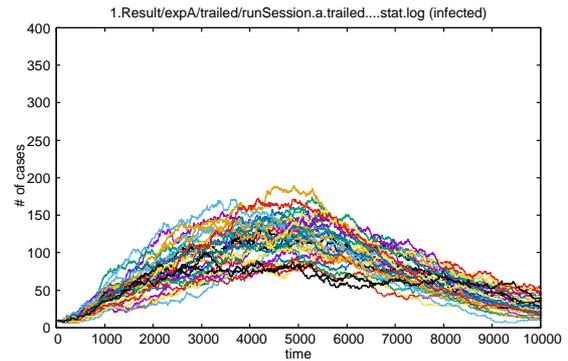


図 3 同一軌跡による感染拡大シミュレーション結果

状態で、活性化した感染状態のエージェントと一定距離 (感染距離) に近づくと、1 サイクルで一定確率 β で感染状態 (不活性) に遷移する。

- 感染状態 (不活性): ウィルスに感染しているが、まだ他者に感染を広げる能力のない状態。この状態のまま一定時間サイクル (活性化遅延時間) 経過すると、活性化する。また、1 サイクルで一定確率 γ で治癒状態に遷移する。
- 感染状態 (活性化): ウィルスに感染し、他者に感染を広げる能力のある状態。また、1 サイクルで一定確率 γ で治癒状態に遷移する。
- 治癒状態: 治癒した状態。免疫があり、2 度めの感染は生じないものとする。

図 1 は、上記の感染状態遷移を表したものである。なお、以下の実験では、感染距離 = 0.3、感染確率 $\beta = 0.15$ 、治癒確率 $\gamma = 0.001$ 、活性化遅延時間 = 100 サイクルとしている。

移動モデルはランダムウォークとする。すなわちエージェントは、各サイクルに於いてランダムに 1 ステップを決め、移動する。1 ステップは、実験では、標準偏差 0.2 の対称な 2 次元正規分布にしたがって決められる。エージェントが動きまわる空間は 100×100 の大きさの正方形であり、各辺はトーラス上に接続し、端がない空間とする。エージェントの総数は 1000 体とし、初期配置は空間全体に渡り一様ランダムで選ばれるとする。また、初期の感染者数は全体の 1% である 10 体とする。

上記の設定でエージェントシミュレーションを行っている様子を図 2 に示す。

2.2 同一軌跡シミュレーション

本稿で示していく実験は、同一軌跡によるシミュレーションである。2.1 節で述べたシミュレーションモデルでは、エージェントの移動過程と感染過程の両方に確率過程が入っているため、どちらに起因する確率的ばらつきかわかりにくいという問題がある。これを避けるために、以下のシミュレーション実験では、一度シミュレ

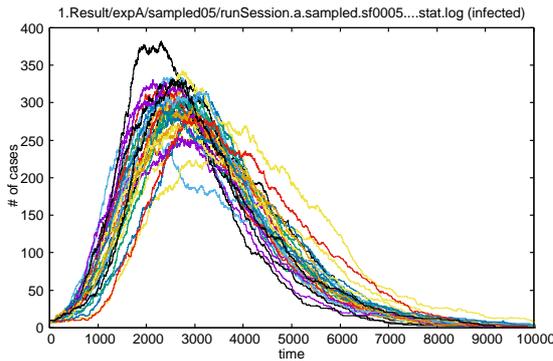


図 4 サンプルングとグリッド化によるシミュレーション結果

シミュレーションを回してエージェント群の軌跡を記録しておき、その軌跡データを固定して、感染過程を複数シミュレーションしていく。これにより、感染過程の確率によるばらつきに集中して議論を行うことができる。

実際に1つの軌跡に対し、異なる初期感染者配置および乱数シードを用いて、32回の拡散シミュレーションを行った結果を図3に示す。このグラフでは、横軸が時間経過、縦軸に感染者数(その時刻における感染状態にあるエージェントの数)を示している。また、各プロット線は32回の個々のシミュレーションの経過を示している。これに示されるように、同じ軌跡を使っても、この程度の感染過程のばらつきがあることがわかる。ただ、そのばらつきの中で、感染の拡大のピークは4000-5000サイクルの間であり、ピークの高さは80-160程度であることがわかる。以下の議論では、この感染の傾向にできるだけ近い結果を再現することが目標となる。

2.3 サンプルングとグリッド化によるシミュレーション

人の移動軌跡に関する実データを用いる場合、サンプルングと粗視化の影響を考慮する必要がある。実際の人々の移動軌跡を示す実データとして現状手に入るものとしては、携帯端末のアプリによる軌跡データがもっとも詳しい。これは、利用許諾のもと、スマートフォンのアプリが収集した位置データを集積したものである。ただし、アプリの利用データであるため、人口カバー率は数%程度であり、プライバシーに配慮して、位置データもグリッド化するなど粗視化されている。よって、このようなデータを使って人の移動に関わるシミュレーションを行う場合、人口カバー率に対応するサンプルングと粗視化のためのグリッドの影響を解析しておく必要がある。

このサンプルングとグリッド化の影響を見るため、2.2節で用いた軌跡データに対し、サンプルングとグリッド化したデータを用いて、もとの規模の移動データに擬似的に復元して感染シミュレーションを行う実験を行った。具体的なシミュレーションは以下の手順で行った。(1) サンプルングとしては、もとのエージェントの20%、

すなわち200体の軌跡のみを代表として抽出する。

- (2) 空間を 10×10 のグリッドに分割し、各軌跡の各時点の位置情報をグリッド番号で代表させるように粗視化する。
- (3) 上記の1のステップでサンプルングされたデータをもとの人口に復元するため、1つの代表軌跡に対して5体分のエージェントを割り当てる。
- (4) 上記2のステップでの粗視化の復元のため、各エージェントの各時刻における位置は、代表軌跡のグリッドに対し、そのグリッド内に一様ランダムに配置する。
- (5) このようにして復元された位置情報を用いて、2.2節と同様に感染シミュレーションを行う。

上記の設定で行った結果を図4に示す。これを見ると、感染の拡大スピードがあがり、早期にかつ大きく感染拡大が生じていることがわかる。すなわち、感染のピーク時期は2000-3000サイクルぐらいに早まり、ピークの高さも250-350程度になっている。

感染拡大が強調されてしまう原因としては以下の2つの原因が考えられる。

- (A) サンプルングの復元のため、同じ位置情報でエージェントを増幅するため、どうしても位置が偏ってしまうと考えられる。感染シミュレーションでは、人が集中すると感染が高確率で発生するため、位置データの偏りは感染拡大を増幅してしまう原因になる。
- (B) 粗視化の復元のためにグリッド内で一様ランダムでエージェント位置を決めていることも感染拡大に寄与していると考えられる。2.1節で述べたエージェントの移動はランダムウォークであり、空間内を連続的に移動する。すなわち、一度接触したエージェント同士はその後も近くを彷徨うため、感染者が短時間で接触できる可能性のある未感染者は局所的に限定される。一方、一様ランダムで位置を変えると、接触できる未感染者はグリッド全域にわたるため、より広範囲のエージェントと接触する可能性が生じるため、感染拡大の確率が増えることになる。これらの2つの要因の影響により、図4に示されるような感染拡大の促進が起きていると考えられる。

3. グリッド内ランダムウォークによる補正

2.3節の議論のうち、まず、(B)で示した要因を緩和する方法について議論する。

前節で指摘しているように、グリッドに粗視化した位置情報を一様ランダムに復元しているのが、感染拡大強調の原因である。そこで、一様ランダムに代わり、各々のエージェントが、グリッド内でランダムウォーク

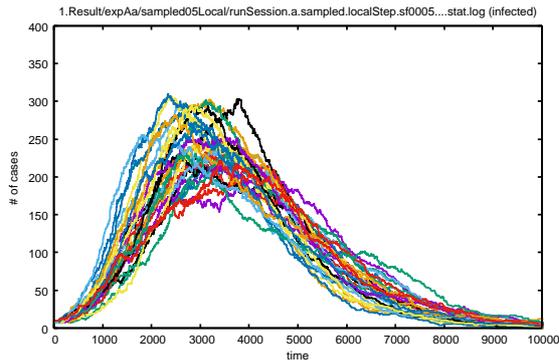


図 5 サンプルング+グリッド内ランダムウォークによるシミュレーション結果

くしていると仮定してみる。すなわち、各エージェントは、グリッドの代表点に対して相対的にランダムウォークを行っているとし、各時刻のエージェントの位置は、サンプルングの代表データで示されるグリッドの中で、その相対的なランダムウォークを足し合わせたものとして扱う。ただし、その相対的なランダムウォークは、境界をトラス状に接続したグリッド内で行われるものとする*1。

そのような設定でこれまでと同様の感染シミュレーションを行った結果を図5に示す。これを見ると、感染のピークの時期が2500–4000サイクル、ピークの高さが200–300となっている。図4と比べると、ピーク時期はおそくなり、高さも抑えられ、図3に近づいていることがわかる。

これにより、グリッド内での位置を、一様ランダムからランダムウォークにすることで、感染拡大の影響を低減させることができていると考えることができる。

4. サンプルング復元による感染拡大の補正

次に、2.3節で議論した(A)の影響の解消について議論する。ここでは、サンプルングしたデータを増幅して復元する際に、どのように統計的性質が変わるかをグリッド毎の人口を中心に見ていく。

4.1 サンプルング後増幅とグリッド内人口の統計値

空間を構成するグリッドの、各グリッド内における人口の確率分布が $\rho(n)$ で与えられ、その平均・分散が各々 μ_n, σ_n^2 とする。

*1 グリッド毎にトラスとなるため、全空間で見ると、グリッド境界付近で、グリッドの反対側に跳躍するエージェントが生じる。また、サンプルングした代表軌跡のグリッドが、時間サイクルで変わっていくと、そこでも跳躍現象が生じる。その意味で、完全なランダムウォークを再現することには成っていない。ただ、同じグリッド内に存在するエージェントが、時間の経過で局所的にしか接触範囲を広げられない、という性質は再現できていると考える。

$$\mu_n = \sum_n n\rho(n) \quad (1)$$

$$\sigma_n^2 = \sum_n (n - \mu_n)^2 \rho(n) \quad (2)$$

人口 n のグリッドを、サンプルング率 $\frac{1}{m}$ (m 人に一人の割合でサンプルング)でサンプルングした時、そのサンプル数 k の分布は、二項分布 $\mathcal{B}(k; n, \frac{1}{m})$ に従うので、人口 n のグリッドのサンプル数 k の平均 $\mu_{k|n}$ ・分散 $\sigma_{k|n}^2$ は以下ようになる。

$$\mu_{k|n} = \frac{n}{m} \quad (3)$$

$$\sigma_{k|n}^2 + \mu_{k|n}^2 = \frac{n(m+n-1)}{m^2} \quad (4)$$

$$\sigma_{k|n}^2 = \frac{n(m-1)}{m^2} \quad (5)$$

グリッド全体に渡る、グリッド毎サンプル数 k の分布 $\mathcal{P}(k)$ を考えると、その分布及び k の平均 μ_k 、分散 σ_k^2 は以下ようになる。

$$\mathcal{P}(k) = \sum_n \mathcal{B}\left(k|n, \frac{1}{m}\right) \rho(n) \quad (6)$$

$$\begin{aligned} \mu_k &= \sum_k k\mathcal{P}(k) \\ &= \frac{\mu_n}{m} \end{aligned} \quad (7)$$

$$\begin{aligned} \sigma_k^2 + \mu_k^2 &= \sum_k k^2\mathcal{P}(k) \\ &= \frac{\sigma_n^2 + \mu_n^2 + (m-1)\mu_n}{m^2} \\ \sigma_k^2 &= \frac{\sigma_n^2 + (m-1)\mu_n}{m^2} \end{aligned} \quad (8)$$

$\frac{1}{m}$ でサンプルングした人口 k を m 倍してもどした擬似人口 $\nu = mk$ とその分布 $\mathcal{P}(\nu)$ を考える。この時、擬似人口 ν の平均 μ_ν 、分散 σ_ν^2 は以下となる。

$$\begin{aligned} \mu_\nu &= \sum_\nu \nu\mathcal{P}(\nu) \\ &= \mu_n \end{aligned} \quad (9)$$

$$\begin{aligned} \sigma_\nu^2 &= \sum_\nu (\nu - \mu_\nu)^2 \mathcal{P}(\nu) \\ &= \sigma_n^2 + (m-1)\mu_n \end{aligned} \quad (10)$$

4.2 感染者率と拡大速度

人口に対する感染者の割合を感染者率、治癒者の割合を治癒者率、とし、以下の記号で表す。

感染者率： λ

治癒者率： θ

つまり、全人口 N に対し、感染者数、治癒者数、未感染者数は以下になる。

感染者数： λN

治癒者数： θN

未感染者数： $(1 - \lambda - \theta)N$

感染拡大は、未感染者が感染者と距離 r 内にいる時に、単位時間あたり、感染率 β で生じるものとする。

面積 s のグリッド内で、ある未感染者がある感染者から距離 r 内にいる確率 q は、

$$q = \frac{\pi r^2}{s} \quad (11)$$

よって、その二人の間で感染拡大が生じる確率は、

$$\mathcal{P}(\text{個々の感染拡大}) = \frac{\beta \pi r^2}{s} \quad (12)$$

そのグリッド内の人口が n 人とすれば、感染者は λn 人いる。よって、ある未感染者が感染させられる確率 ψ は、

$$\begin{aligned} \psi &= 1 - \left(1 - \frac{\beta \pi r^2}{s}\right)^{\lambda n} \\ &\sim 1 - \left(1 - (\lambda n) \frac{\beta \pi r^2}{s}\right) \\ &= (\lambda n) \frac{\beta \pi r^2}{s} \end{aligned} \quad (13)$$

人口 n のグリッドでの患者数の増加数 $\Delta_{\lambda n}$ は、

$$\begin{aligned} \Delta_{\lambda n} &= (1 - \lambda - \theta)n\psi \\ &= (1 - \lambda - \theta)\lambda n^2 \frac{\beta \pi r^2}{s} \end{aligned} \quad (14)$$

よって、グリッド内の増加数の平均 $|\Delta_{\lambda n}|$ は、

$$\begin{aligned} |\Delta_{\lambda n}| &= \sum_n \mathcal{P}(n) \Delta_{\lambda n} \\ &= \sum_n \mathcal{P}(n) n^2 (1 - \lambda - \theta) \lambda \frac{\beta \pi r^2}{s} \\ &= (\mu_n^2 + \sigma_n^2) (1 - \lambda - \theta) \lambda \frac{\beta \pi r^2}{s} \end{aligned} \quad (15)$$

一方、 $(1/m)$ でサンプルしたものを m 倍して復元した場合には、グリッド内の擬似人口における増加数 $\Delta_{\lambda \nu}$ の平均は、

$$\begin{aligned} |\Delta_{\lambda \nu}| &= (\mu_\nu^2 + \sigma_\nu^2) (1 - \lambda - \theta) \lambda \frac{\beta \pi r^2}{s} \\ &= (\mu_n^2 + \sigma_n^2 + (m-1)\mu_n) \\ &\quad \cdot (1 - \lambda - \theta) \lambda \frac{\beta \pi r^2}{s} \end{aligned} \quad (16)$$

15 式と 16 式を比較すると、 $1/m$ サンプルングおよびグリッド内での m 倍により、感染拡大速度は、 $(\mu_n^2 + \sigma_n^2 + (m-1)\mu_n)/(\mu_n^2 + \sigma_n^2)$ 倍される。これを逆に適応すると、サンプルング後に増幅する場合、感染率 β を $(\mu_n^2 + \sigma_n^2)/(\mu_n^2 + \sigma_n^2 + (m-1)\mu_n)$ 倍して補正する必要がある。

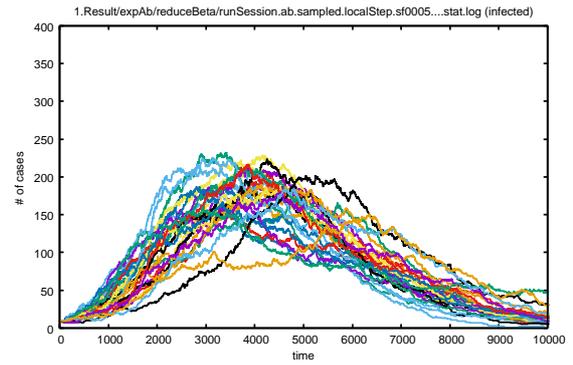


図 6 $\beta' = 0.11$ に補正した時の結果

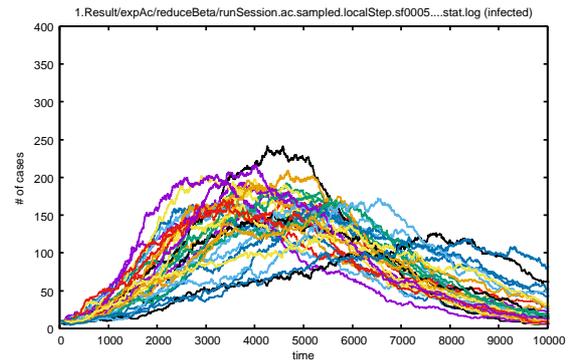


図 7 $\beta' = 0.10$ に補正した時の結果

表 1 各シミュレーションにおける感染ピーク

設定	ピーク時期	高さ
original	4000~5000	80~160
sample & grid	2000~3000	250~350
in-grid random walk	2500~4000	200~300
+ adjusted β	3000~5000	120~220
+ additional adjusted β	3500~6000	120~200

4.3 実験による検証

4.2 節で求めたように、感染率 β を $(\mu_n^2 + \sigma_n^2)/(\mu_n^2 + \sigma_n^2 + (m-1)\mu_n)$ 倍に補正してシミュレーションを行う。本実験の場合、補正した感染率は $\beta' = 0.11$ となる。なお、3 節で提案した補正方法もこの実験では取り込んでいる。シミュレーション結果を図 6 に示す。この場合、感染拡大のピーク時期は 3000~5000 サイクルとなり、ピークの高さは 120~220 程度になっている。これまでの感染ピークの傾向をまとめると、表 1 となる。これを見るとわかるように、図 4 や図 5 に比べ、3 節の感染拡大過程にだいぶ近づいていることがわかる。

5. 最後に

本稿では、人の移動軌跡データにより感染シミュレーションにおいて内包されるサンプルングとグリッド粗視化の影響について、その影響の解消を行うための手法を 2 つ、導出し、実験によりその効果を確認した。

グリッド粗視化の影響の解消では、エージェントの移

動モデルを局所的に適用することで、粗視化された位置データを単純にランダムに復元するのに比べ、感染拡大の影響を軽減する方法を提案した。具体的な移動モデルはランダムウォークモデルを仮定し、提案手法の効果を実験により確認した。

また、サンプリングとその増幅の影響については、人口の分散値の違いを導き出し、それが感染拡大に与える影響から、感染率を補正する方法を導出した。この方法を、上記のグリッド粗視化に対する方法と組み合わせて実験し、感染拡大強調の影響を軽減する効果を確認した。

これらの方法で、サンプリングとグリッド粗視化されたデータの素朴な復元によるシミュレーションへの影響を、かなり軽減できることが実験で示された。

ただ、求めた方法ではまだ3節に比べて感染拡大傾向が強調されているように見える。これは、SIRモデルに代表される感染拡大の過程が時間的に指数的な過程であり、僅かな違いが加速的に積分され、どちらかというところの正の方向に強調されがちであるためとも考えられる。比較として、感染率の補正をより強くして $\beta' = 0.10$ とした場合の結果を図7に示す。こちらはだいぶ図3の感染拡大過程に近くなっていることがわかる。

今後は、さらなる補正方法を求めていくとともに、ランダムウォーク以外の移動モデルへの拡張、より大きな系や現実の空間データでの検証を進めていく必要がある。

謝辞 本研究の一部はJSPS科研費18H03301の助成を受けたものです。

参考文献

- [1] Rockett, R. J., Arnott, A., Lam, C., Sadsad, R., Timms, V., Gray, K.-A., Eden, J.-S., Chang, S., Gall, M., Draper, J., Sim, E. M., Bachmann, N. L., Carter, I., Basile, K., Byun, R., O'Sullivan, M. V., Chen, S. C.-A., Maddocks, S., Sorrell, T. C., Dwyer, D. E., Holmes, E. C., Kok, J., Prokopenko, M. and Sintchenko, V.: Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling, *Nature Medicine*, pp. (online) (online), DOI: <https://doi.org/10.1038/s41591-020-1000-7> (2020). Letter published: 09 July 2020.
- [2] Cuevas, E.: An agent-based model to evaluate the COVID-19 transmission risks in facilities, *Comput. Biol. Med.*, Vol. 121, p. 103827 (online), DOI: [10.1016/j.combiomed.2020.103827](https://doi.org/10.1016/j.combiomed.2020.103827) (2020). Published online 2020 May 20.
- [3] Gharakhanlou, N. M. and Hooshangi, N.: Spatio-temporal simulation of the novel coronavirus (COVID-19) outbreak using the agent-based modeling approach (case study: Urmia, Iran), *Inform. Med. Unlocked*, Vol. 20, p. 100403 (online), DOI: [10.1016/j.imu.2020.100403](https://doi.org/10.1016/j.imu.2020.100403) (2020). Published online 2020 Jul 30.