

大規模イベントにおける疑似観測データを用いた観測データ量の十分性評価

Assessing the sufficiency of the amount of observation data using pseudo-observation data in large-scale events

櫻井 舜¹⁾ 重中 秀介²⁾ 山下 倫央¹⁾ 大西 正輝²⁾
Sakurai Shun Shigenaka Shusuke Tomohisa Yamashita Masaki Onishi

1 序論

2001 年の明石花火大会歩道橋事故や 2015 年のメナー群集事故, 2022 年のソウル梨泰院雑踏事故など, 大勢の人が集まるイベントでは混雑による群集事故が問題となっている. 事故原因を特定し人流を適切に誘導することが群集事故の未然防止につながる [1]. カメラや GPS などの計測機器は, 過密エリアの監視や人流の追跡に役立つため, 群集事故の特定に有用である. しかし, 群集事故は様々な人の意思決定が複雑に絡み合った結果で発生するため, 計測機器のみから適切な人流誘導を策定するのは困難である. そのため近年, 人の意思決定をモデルとして加える群集シミュレーションが, 人流誘導の策定に用いられている [2].

現実世界の様々な現象を仮想空間上に再現し, 一致させるデータ同化は, シミュレーションの更なる性能向上に役立つ方法として期待されている. データ同化は, 現実世界で収集した観測データを基に, シミュレーションのモデルやパラメータを調整する作業である. イベント終了時の帰宅や災害時の避難では, シミュレーションによる人流全体の状況把握にデータ同化がよく用いられる. しかし, 数千から数万人規模の大規模イベントでは, 計測機器の費用や個人のプライバシーなどの問題で, 十分な量の観測データが得られないことがある. したがって, データ同化に用いる観測データが人流全体の再現に十分な量であるかを評価する必要がある. 上記が達成されると, シミュレーションの信頼性が保証されるだけでなく, 機器費用やアルバイトの費用の削減などのコスト効率と持続可能性の向上を基づいた利益を得ることができる.

本研究が議論する観測データ量の十分性は, 多様な分野での標本サイズ決定に関する重要な議論であり, 研究の信頼性と妥当性を確保するために不可欠である. 従来の標本サイズ決定方法として, z スコアや許容誤差などの統計式が広く活用されている. z スコアは推定値が真の値をとる確率を示す値であり, 許容誤差は推定値を真の値と許容する幅である. これらの統計式はデータが独立性や等分散性を持つといった仮定を置くため, 統計的な偏りを最小限に抑え, 信頼性の高い評価が可能である. 主に医療や社会科学の分野では古くから適用されている.

しかし, 最低限必要な標本サイズはデータの特徴に依存するため, 統計的な数値を幅広い分野で一律に適用することは困難である [3]. 特に理論的な前提条件を満たさないデータでは, 様々な値に偏りが生じる. 具体的には, 等分散性が満たされない場合, 標準誤差が上方に偏ることがわかっている [4]. 実際, 本研究で扱う人の移

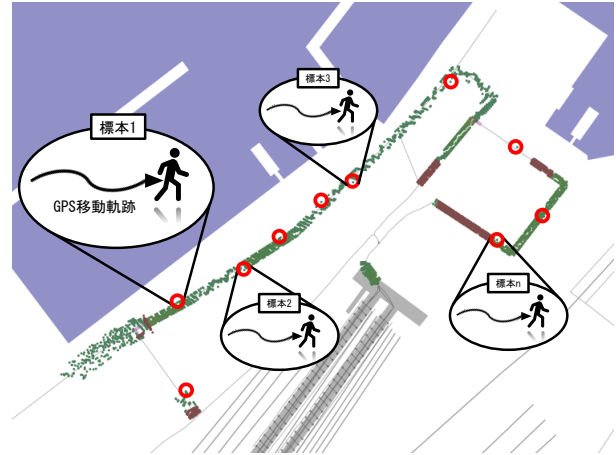


図 1 シミュレーションを用いてイベントを再現し, 全参加者を母集団, 計測される一部参加者を標本とする

動軌跡は時間経過により連続的に変化することから等分散性や独立性のような理論的な前提条件を完全には満たすことができないと予想される. そのため, 統計式のみで観測データ量の十分性を厳密に評価することは困難だと考える.

そこで, シミュレーションを用いた標本サイズの評価手法を提案する. まず図 1 のようにシミュレーションを用いて仮想的に母集団を生成し, 標本を取得する. 標本サイズごとに統計的な手法を用いた理論値と提案手法を用いた実験値の両方を算出し, 評価を行う.

本稿の構成は以下の通りである. 2 節では従来の標本サイズ決定手法を紹介する. 3 節でデータや観測, 4 節で実験設定の説明, 5 節で実験の結果について述べる. そして最後に, 6 節で結論と今後の展望を述べる.

2 従来手法

本節では, 標本サイズを導出する統計式の 1 つについて述べる. 次の式は, アンケート調査における標本サイズの導出に広く利用される統計式である.

$$n = \frac{z^2 p(1-p)}{1 + \frac{e^2}{z^2 p(1-p)} e^2 N} \quad (1)$$

ここで, n は最低限必要な標本サイズ, p は確率, N は母集団の規模, z は z スコア, e は許容誤差を示す.

以降, 変数の詳細について述べる. 母集団は, 対象となる集団の全体を表す. 標本は, 母集団から無作為に選択された一部の要素であり, 母集団を表現することを目的としている. そのため, 母集団の大きさと最低限必要となる標本のサイズは比例する. 確率は, 試行の成功率や事象の期待値を表す. アンケート調査の場合, 特定の回答をする確率や特定の属性を持つ人の割合などが該当する. z スコアは標準正規分布の分位点を表し, 特定の

1) 北海道大学 Hokkaido University

2) 産業技術総合研究所 National Institute of Advanced Industrial Science and Technology



図 2 関門海峡花火大会の帰宅経路

信頼区間に対応する値を使用する。信頼区間は統計学において、推定値の信頼性を示すために使用され、値が後述の許容誤差の中にある確率を示す区間である。許容誤差は、調査や統計的推定の信頼性を評価する指標であり、信頼区間の半径として表現される。一般に、5%許容誤差と 95% 信頼区間に対応する z スコアが用いられる。

統計式は正規分布などの一般的な分布を想定しているため、偏りのあるデータや特徴のあるデータを扱う場合は標本サイズの推定が困難である。本研究で扱う人の移動軌跡は、偏りのあるデータと考えられるため、実験的な評価が必要である。

3 準備

西日本最大規模の関門海峡花火大会では、帰宅時の混雑緩和を目的としてイベント終了時に会場から門司港駅に向かう人の流れを計測する実験を行っている [5, 6]。イベント当日の門司港駅周辺では、車両の乗り入れが禁止されているため、参加者の多くは徒歩での移動を強いられる。参加者の移動データは、帰宅経路にハンディ GPS を持った被験者を発生させ、GPS を計測することによって記録される。GPS は衛星から送られてくる電波を受信することで自分自身のグローバルな位置情報を知ることができるシステムであり、デバイスを持った人の位置座標を秒単位で取得することができる。本研究では、GPS 計測によって得られた帰宅時の参加者の移動軌跡を標本として用いる。

花火大会では、イベント会場から門司港駅に向かう帰宅者を 3 つの帰宅経路に分けて誘導している。3 つの帰宅経路は図 2 に示した通りである。それぞれの経路上には警備員が配置されており、警備員はルートを変更する分岐制御と歩行者の動きを止める停止制御を行う。最も短い経路 (320 m) に配置されている警備員の数は分岐制御 1 人、停止制御 3 人である。次に短い経路 (600 m) に配置されている警備員の数は分岐制御 2 人、停止制御 2 人である。最も長い経路 (770 m) に配置されている警備員の数は分岐制御 2 人、停止制御 2 人である。各エージェントの帰宅経路は警備員の指示によって決定する。

人の流れの計算には、オープンソースのマルチエージェント型の歩行者シミュレータ CrowdWalk¹⁾ [7] を用いる。これは、一次元歩行シミュレータであり、帰宅時

の人の流れを簡易的に表現することで、二次元の歩行者シミュレータよりも速い計算を可能にする。さらに、CrowdWalk は Social Force model [8] に基づき加速度を決定するため、精密な歩行者の移動の計算や高い再現精度が期待できる。このシミュレーションは、実際の避難計測とシミュレーション結果の比較が行われており、二次元シミュレータとほとんど変わらない結果が得られている [9]。

本研究では、シミュレーション内での歩行者の挙動を次の 2 タイプに分ける。

- **Minor Agent:** GPS で計測されていない歩行者である。分岐制御と停止制御に従うため、全ての経路で発生する可能性がある。
- **Measured Agent:** GPS で計測される歩行者である。分岐信号に従わず、あらかじめ定められた経路の停止制御に従う。

人流データは、単位時間当たりの歩行者の緯度経度情報を距離情報に変換したデータである。本研究が扱う群集シミュレーションでは、Measured Agent が 1 秒当たりの緯度経度情報を取得し、これを距離情報に変換した移動軌跡を人流データとして扱う。

4 評価手法

本研究では、シミュレーションによる人流全体の状況把握を目的として、観測人数の十分性を評価する。評価手順は次の通りである。まず、正解の人流データ $D = \{d_1, \dots, d_N\}$ を取得する。次に、無作為に作成されたパラメータによって B 個の人流データ $\{\hat{D}_1, \dots, \hat{D}_B\}$ を取得する。この時、各パラメータの人流データは $\hat{D}_b = \{\hat{d}_{b,1}, \dots, \hat{d}_{b,N}\}$ である。次に、正解とランダムな観測人数別の人流データを実験的手法と理論的手法それぞれを用いて評価する。最後に、相関係数の推移と理論式を重ね合わせ、比較を行い、観測人数の十分性を評価する。ただし、本論文で扱う正解の人流データは、過去に実施されたイベントの観測データに基づいたシミュレーションを用いて作成した合成データである。

4.1 実験的評価と理論的評価

本節では、実験的手法および理論的手法を用いた観測人数の十分性評価手順について説明する。

実験的方法の手順としては正解とランダムな人流データを取得した後、下記の式 (2) を用いて 2 つの移動軌跡の Root Mean Square Error (RMSE) を計算する。

$$f(b, x, r, i) = \frac{x}{N'} \sum \sqrt{\sum_{t=0}^{N'} \frac{\max(\hat{T}_{b,i}, T_i) (d_i(t) - \hat{d}_{b,i}(t))^2}{\max(\hat{T}_{b,i}, T_i) L_r}} \quad (2)$$

ここで、 i は出発時刻を示し、19:00~24:00 までの間の 1 分刻みの値を取り、 R は経路数、 L_r は経路 $r \in \{1, \dots, R\}$ の総距離、 N' は $\frac{N}{R}$ で求められ、各経路の母集団の規模を表す。 $d(t)$ は出発時刻から t 分経過した時の累計移動距離、 T_i は出発時刻 i の Measured Agent が駅に到着する時刻、 x はハンディ GPS を持った被験者の出発間隔を示す。本研究では、ランダムパラメータ数は $B = 1000$ 、母集団の規模は $N = 900$ 、経路数は $R = 3$ 、各経路の総距離は $L_r \in \{320, 600, 770\}$ 、GPS の取得間隔は $x \in \{1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 25, 30, 50, 60, 75, 100, 150, 300\}$ とする。

1) <https://github.com/crest-cassia/CrowdWalk>

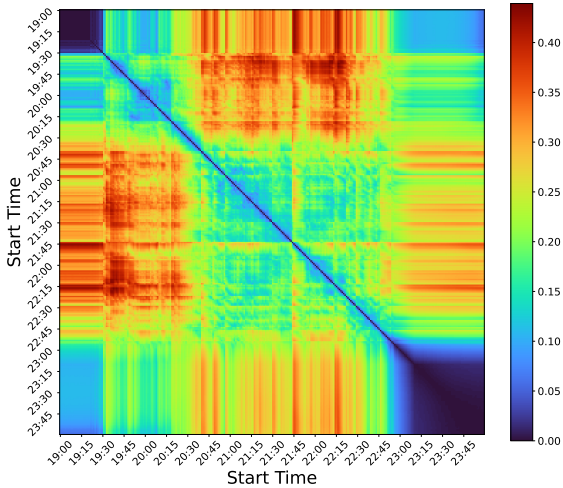


図 3 出発時刻の差と移動軌跡の RMSE の関係を表すヒートマップ

これらの計算結果を用いて、母集団と標本の計算結果の相関係数を求める。ここで求めた相関係数は母集団との類似度を表現しているため、許容誤差と近い意味を持つと考え、実験の評価に用いる。

理論的方法は式 (1) を変形した以下の式を使用する。

$$e = z \sqrt{\frac{p(1-p)(N-n)}{nN}} \quad (3)$$

$$y = 1 - e \quad (4)$$

$$y = 1 - 1.96 \sqrt{\frac{p(1-p)(Rx-1)}{N}} \quad (5)$$

ここで、 n , p , N , z , e は式 (1) と同様の変数、 y は正解率を示す。本研究では、データの特徴の分析により p を算出し、信頼区間は 95% に設定するため z スコアは 1.96 とする。

z スコア、母集団の規模、確率、標本サイズを入力として許容誤差を求め、1 から引くことで正解率を出力している。

5 実験

5.1 人流データの特徴分析

出発時刻と移動軌跡の関係性を可視化し、データの特徴を分析した。特徴分析には、2015 年から 2018 年の計 4 年分の正解データを用いる。同年の同経路を移動する Measured Agent の移動軌跡の RMSE を下記の式 (6) を用いて計算し、出発時刻が移動軌跡に与える影響の程度を数値化した。

$$g(r, i, j) = \sqrt{\frac{\sum_{t=0}^{\max(T_i, T_j)} (d_i(t) - d_j(t))^2}{\max(T_i, T_j) L_r}} \quad (6)$$

ここで、 j は i と同様に出発時刻を示し、 R , L_r , $d(t)$, T_i は式 (2) と同様の変数を示す。

19:00～24:00 までの間、1 分刻みで発生する Measured Agent の移動軌跡を取得することから、式 (6) の計算により 300×300 のデータを計 12 個 (3 経路 \times 4 年分) 作成した。それら計 12 個のデータの平均を取り、ヒートマップを用いて図示したものが図 3 である。この図から、人流データが持っている 2 種類の特徴が読み取れる。1 つ目は、出発時刻の差が小さいとき、移動軌跡の

表 1 Guilford rule

Correlation coefficient (ρ)	Strength of relationship
$0.00 \leq \rho \leq 0.20$	Negligible
$0.20 < \rho \leq 0.40$	Weak
$0.40 < \rho \leq 0.70$	Moderate
$0.70 < \rho \leq 0.90$	Strong
$0.90 < \rho \leq 1.00$	Very Strong

差も小さく、データの特徴が酷似しているということである。これは移動軌跡が時系的に依存することを示すため、人流データは独立性を持たないと言える。2 つ目は、20:40～22:20 の間に出発した Agent の移動軌跡とそれ以外の時刻に出発した Agent の移動軌跡は RMSE が大きく、データの特徴が大きく異なるということである。これは出発時刻に依存した偏りが存在することを示すため、等分散性を持たないと言える。これらのことから、統計式を適用する際の理論的な前提条件を人流データには満たすことができないと考えられる。

5.2 理論値と実験値の比較

ここでは、データの特徴に基づいた標本サイズの実験的な評価と理論的な評価の比較を行う。図 4 は、類似度から導出した実験値と許容誤差から導出した理論値を図示したものである。x 軸は、ハンディ GPS を持った被験者の出発間隔を示す。y 軸は、理論値の場合、式 (5) で求めた正解率を示し、実験値の場合、母集団と標本の相関係数を示す。図から、x 軸方向に 3 つに分割され、それぞれ異なるデータの特徴が現れていることがわかる。

GPS 取得間隔が 1～20 分の時、実験値の概形は x 軸に沿って平行であり、一方で理論値は大幅に負の傾きを持つことがわかる。図 3 では、出発時刻の差が小さい場合、移動軌跡の誤差も小さい傾向がある。また、GPS を取得していない時刻の情報は取得した時刻の情報と酷似していたと言える。したがって、取得した情報のみで人流全体を表現することが十分可能であると考えられる。

次に GPS 取得間隔が 20～100 分の時、実験値と理論値の傾きは近い値を示すことがわかる。これは取得間隔の増加に伴い、GPS を取得していない情報と取得した情報の類似度が徐々に減少したことが原因と考えられる。

最後に GPS 取得間隔が 100 分よりも大きい時、理論値よりも実験値の方が大きい負の傾きを持つことがわかる。人流データの特徴分析から 20:40～22:20 の間の 100 分間とそれ以外のデータは大きく特徴が異なり、データの特徴によって 2 つに分類できるとわかっている。そして、100 分よりも GPS 取得間隔が大きくなることで、2 種類両方の情報を取得できない可能性が大きくなる。よって、人流全体の把握が難しくなり、大きい負の傾きを持つと考えられる。

これらのことから、実験値は人流データの特徴を捉えていることがわかる。

5.3 標本サイズの十分性評価

実験的手法はデータの特徴を捉えることができるため、十分性の評価において適していると考えられる。実験値を用いて標本サイズを評価する方法は 2 種類存在する。

1 つ目は、実験値を相関係数として扱い、ギルフォードルールを適用する方法である [10]。ギルフォードルールは、相関係数の値を相関の強さで分類する指標であり、表 1 に示した通りである。理論値が 0.90 よりも大

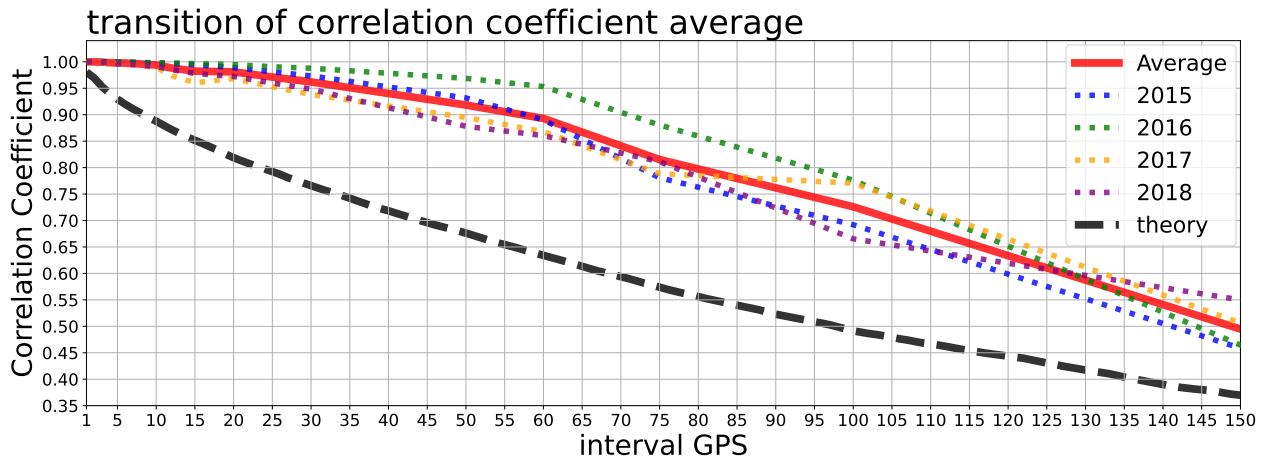


図4 理論値と実験値

きい GPS 取得間隔は 1~8 分程度である。実験値が 0.90 よりも大きい GPS 取得間隔は 1~40 分程度である。ギルフォードルールを指標として用いる場合、理論値と実験値の間には約 5 倍もの差があることがわかる。

2 つ目は、実験値を許容誤差として扱い、一般的に用いられる指標を適用する方法である。許容誤差は一般的に 5%とされている。理論値が 0.95 よりも大きい GPS 取得間隔は 1~2 分程度である。実験値が 0.95 よりも大きい GPS 取得間隔は 1~25 分程度である。一般的に用いられる指標を用いる場合、理論値と実験値の間には約 12 倍もの差があることがわかる。両方の指標においても十分性が確保されている必要があると考えるため、これら二つの指標の共通の範囲から理論値の場合は取得間隔 2 分、実験値の場合は取得間隔 25 分を十分量とする。

6 結論

本研究では、統計式による理論的評価値と提案手法による実験的評価値を比較することで標本サイズの十分性を検証した。その結果、統計式を用いて導出される標本サイズの理論値は実際に最低限必要となる標本サイズよりも大きい値を示した。これは、人の移動軌跡が持つ時間的な偏りは統計式に反映できないことを意味する。つまり、本研究における標本数の十分性は実験的評価値が適していると考えられる。この標本サイズの評価を通じて得られる利益について検討を行う。許容誤差を 5%を境界とした場合、統計的手法を用いた理論的評価では、GPS 数は 2 分刻みなので 450 台が必要とされる。提案手法を用いた実験的評価では 25 分刻みなので 36 台が必要とされる。これらより、実験的評価を行うことで GPS デバイス費用の 92%削減が可能であり、コスト効率の大幅な向上が期待される。

今後の課題を 3 つ提示する。1 つ目は、シミュレーション内の行動ルールを複数作成し十分性評価への寄与度を調査することである。今回の実験ではシミュレーション内の人々は Minor Agent と Measured Agent の 2 種類の行動ルールに集約している。より現実に近いシミュレーションを行うには人ごとに異なる行動ルールを定義する必要がある。2 つ目は、統計的な指標を用いる際の基準値を評価することである。この基準値は特定の状況に依存するものが多い。本研究で扱ったギルフォードルールや推奨許容誤差の 5%、推奨信頼区間の 95%など

が例として挙げられる。これらの基準値の妥当性を評価し、信憑性の高い値を提案する。3 つ目は、シミュレーションの種類を増やして実験を行うことである。本研究では、門司港花火大会のシミュレーションのみを用いたので、経路数や規模、地域などの条件が全て同じであった。本研究の結果がどの条件に依存するのかを調べ、実験的評価手法の汎化性能を調査する必要がある。

参考文献

- [1] G Keith Still. *Introduction to crowd science*. CRC Press, 2014.
- [2] Mohamed H. Dridi. *Simulation of high density pedestrian flow: Microscopic model*, 2015.
- [3] Susan Thorpe Terry Young Konstantina Vasileiou, Julie Barnett. *Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period*. *BMC Medical Research Methodology volume*, 2018.
- [4] Ian A Silver. *The violating assumptions series: Simulated demonstrations to illustrate how assumptions can affect statistical estimates*. 2021.
- [5] 山下倫央, 岩田敦之, 宮部真衣, 荒牧英治. It を活用した群集流動における安全安心の実現に関する実証的研究. 日工組社会安全研究財団 2015 年度一般研究助成研究報告書, 2015.
- [6] 山下倫央, 松島裕康, 野田五十樹. 屋外大規模イベントの流動制御に向けた動的群集誘導モデルの構築. 第 30 回 人工知能学会全国大会論文集, p. 1H5OS05b2, 2016.
- [7] Tomohisa Yamashita, Takashi Okada, and Itsuki Noda. *Implementation of simulation environment for exhaustive analysis of huge-scale pedestrian flow*. *SICE Journal of Control, Measurement, and System Integration*, Vol. 6, No. 2, pp. 137-146, Jan 2013.
- [8] Dirk Helbing and Peter Molnar. *Social force model for pedestrian dynamics*. *Physical review E*, Vol. 51, No. 5, pp. 4282-4286, 1995.
- [9] 山下倫央, 副田俊介, 大西正輝, 依田育士, 野田五十樹ほか. 一次元歩行者モデルを用いた高速避難シミュレータの開発とその応用. *情報処理学会論文誌*, Vol. 53, No. 7, pp. 1732-1744, 2012.
- [10] Joy Paul Guilford. *Fundamental statistics in psychology and education*. 1950. McGraw-Hill.