

混雑統計データを用いたイベント周辺人口予測 のための空間自己回帰モデルの構築

坂 匠¹・薄井 智貴²・山本 俊行³・森川 高行⁴

¹学生会員 名古屋大学大学院 工学研究科 社会基盤工学専攻 (〒464-8603 愛知県名古屋市千種区不老町)

E-mail: ban.takumi@j.mbox.nagoya-u.ac.jp

²正会員 名古屋大学大学院 特任准教授 経済学研究科

E-mail: tomo.usui@nagoya-u.jp

³正会員 名古屋大学 教授 エコトピア科学研究所

E-mail: yamamoto@civil.nagoya-u.ac.jp

⁴正会員 名古屋大学 教授 未来社会創造機構

E-mail: morikawa@nagoya-u.jp

近年、人々の流動を予測する動きが高まってきている。また、携帯電話の普及率の増加に伴いビッグデータの有意性が高まってきた一方、ビッグデータを用いて人々の動きを把握する研究は少ない。本研究では、ユーザーの操作が不要な携帯電話 GPS データを加工しメッシュ毎に集計された「混雑統計®」データを用いて、イベント時における人々の動きを空間同時自己回帰モデルによりモデル化を行った。その結果として、このモデルの AIC は重回帰モデルよりも下がり、当てはまりが良いモデルであると言えた。また、構築したモデルを適用することにより、大体ではあるが混雑区域の推定することが可能である。

Key Words: GPS, "Konzatsu Tokei®", SAR model

1. はじめに

(1) 研究背景と目的

今日の交通行動の特性の一つとして、コンサートなどのイベント開催等の要因により突発的な交通需要が発生し、平常時とは異なる人の流動が挙げられる。それに対してイベントの主催者や参加者、警察等が十分な対応を行うことができないために、人や自動車等の渋滞が発生することもよくみられる。2001 年の明石市民夏まつりにおいて発生した歩道橋事故では、会場の立地条件や警備との連携不足などにより、ボトルネックとなっていた歩道橋にて異常なまでの混雑になり群集雪崩が発生し、多くの死傷者が出てしまう事態となったことは記憶に新しい¹⁾。このような危険を回避するためにも、突発的な交通需要により生じる人々の流動を把握することは重要である。

一方、携帯電話の加入数が急激に増加しており、総務省調べの携帯電話と PHS の加入数の年別推移²⁾によると、現在では 1 人で 2 台所持している場合もあり、普及率は 110%となっている。さらに、2007 年以降に発売された携帯電話には、人工衛星から発信されている電波を受信することにより端末の位置情報を取得することができる GPS(Global Positioning System) 機能が標準で搭載されており、携帯電話利用者の位置情報を送信・蓄積することも可能になってきた。

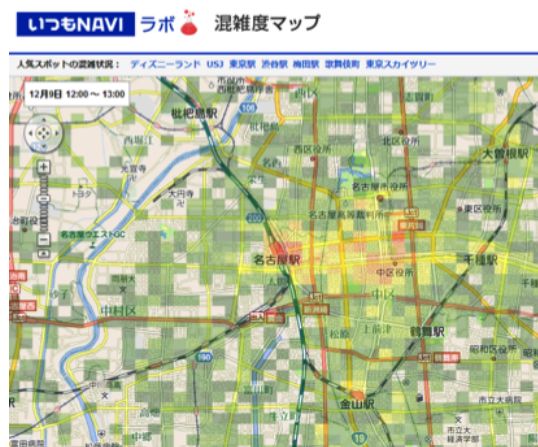


図-1 混雑度マップ ((株) ゼンリンデータコム,2014)

この大規模かつ大容量の GPS データを用いた事例の一つとして、図-1 に示す、株式会社ゼンリンデータコムの「混雑度マップ」³⁾がある。これは、蓄積された GPS データを日本の総人口に当てはめ、メッシュ毎に集計を行い、地図上で可視化したものである。近年になって、このデータのソースとなっている「混雑統計®」データを用いて人の動きを可視化する研究も進められているが、その利用可能性はまだ未知数である。

そこで、本研究では、イベント時の突発的な交通需要を把握することを目的とし、過去におけるイベント時

の人々の位置情報が記録された株式会社ゼンリンデータコム作成の「混雑統計®」データを分析対象とし、イベント時における周辺の流動人口を予測するための、空間同時自己回帰モデルを構築し、検証を行った。

(2) 既往研究と本研究の位置づけ

イベント時の流動人口を予測する研究は多数行われている。羽藤ら⁴⁾は、あるイベント参加者へ PHS を送付し、それにより得られた詳細な位置情報データに対して分析を行った。イベントへの参加を目的とした行動を共にしているグループの構成員により、イベント前後の移動パターンに違いがあることや、イベント時間が近づくにつれイベント会場付近に人々が集中し、イベントの終了時刻を過ぎると急激に人の移動が進むことを明らかにした。イベント時における人々の動きを予測する方法についてもさまざまな手法が提案されている。

一方、「混雑統計®」データを取り扱った研究はまだ少ない。Hayano・Adachi⁵⁾は、250m メッシュ毎に集計された「混雑統計®」データを用いて、福島第一原発事故直後における1時間毎の人の動きを分析した。これまで、福島第一原発周辺の住民が取った避難行動については、避難住民への聞き取りやアンケート調査しか方法がなかった。そこで、「混雑統計®」データといった客観的なデータを用いることで避難行動を可視化し、周辺地域の人数分布の推定を行った。

それらを踏まえた上で、本研究では250m メッシュ毎に統計処理が施された「混雑統計®」データを用いて、過去に開催されたイベント時のデータに対して特徴を抽出し、空間同時自己回帰モデルにより会場周辺の周辺流動人口の推定を行うためのモデルを構築した。

(3) 本論文の構成

本論文は5章で構成される。第1章(本章)では研究背景、研究目的、既往研究と本研究の位置づけについて述べた。第2章では、本研究で利用したデータの概要を示す。第3章では、データの基礎集計結果を示し、考察を行う。第4章では、提案手法による、周辺人口予測モデルの結果を示し、考察を行う。第5章では、得られた知見と今後の課題について述べる。

2. 利用データの概要

本研究では、前述の通り、株式会社ゼンリンデータコムの「混雑統計®」データを利用する。本章では、まずデータの概要と基礎集計結果について述べる。

表-1 データ概要^{6),7)}

取得データ	年月日/時間帯(5分毎)/ メッシュコード/推計人数	
取得日	2012/07/07 (土曜日)	2012/07/13 (金曜日)
取得間隔	5分毎	
取得時間	12:00 - 23:55	
集計メッシュ	250m 四方の5次メッシュ	
イベント名称	プロ野球 (中日-DeNA 戦)	プロ野球 (中日-巨人戦)
開催場所	ナゴヤドーム	ナゴヤドーム
開催時間	14:00 - 17:04	18:00 - 21:57
観客数(人)	約3万5千	約3万8千

(1) 「混雑統計®」データの概要

「混雑統計®」データは docomo の携帯端末の「AUTO GPS」機能を利用して取得したデータで、2009年の冬期から導入され、利用者の許諾を得た上で5分毎に端末の位置情報をサーバーに自動的に送信・蓄積を行っている。特に利用者に操作の必要はなく全自動で行われているため、純粋な位置情報のみのデータとなっている。そのため、携帯電話端末使用者の移動手段を知ることとはできない。また、GPSによる位置情報を取得できない場合は、携帯電話の基地局による位置情報にて補完を行っている。このように取得された位置情報データを株式会社ゼンリンデータコムが独自に拡大係数を設定し、メッシュ毎に統計処理を行ったものが、「混雑統計®」データである。なお、集計する際に使用された拡大係数は未公開となっており、ユーザー個人を特定することは不可能となっている。

(2) 利用データ概要

本研究で利用したデータの概要と、利用データを取得した日に開催されたイベントの詳細を表-1に示す。データ取得範囲は名古屋市のほぼ全域で、5次メッシュで約2,000メッシュとなっている。データの取得間隔は「AUTO GPS」機能の自動送信間隔と同様、5分毎となっている。250m 四方の5次メッシュを用いて集計をされているのは、個人情報保護やGPSの誤差をなくすためである。

対象となるイベントを、表-1に示す。7日と13日にナゴヤドームにて開催されたプロ野球の試合は、7日は試合時間が14時からのデーゲームであり、13日は試合時間が18時からのナイターゲームである。ただし、5分毎に集計されている利用データの仕様上、試合時間をそれぞれ14:00~17:00、18:00~21:55と仮定した。

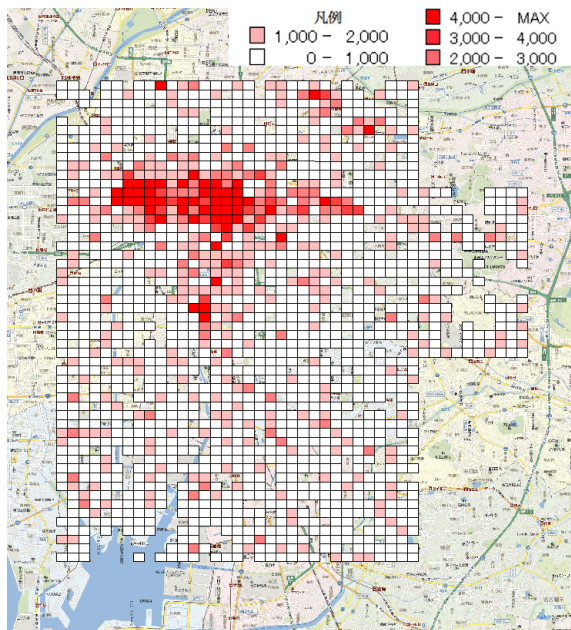


図-2 7/13(20:00)

3. 基礎集計分析

イベント時の人の流動の特徴と傾向を把握するために、まず、利用データの基礎集計の結果について述べる。

(1) 全範囲における基礎集計分析

人の移動傾向を把握するために、GIS 上にて利用データの可視化を行った。13日のイベント開始時刻である18:00から十分に時間が経っている20:00のデータを選択した。図-2は、13日の20:00のデータを示している。地図上のメッシュの色の赤が濃いほどメッシュ内の推計人数が多いことを示している。図中の空白のメッシュはデータが記録されておらず、そのメッシュ内の推計人数は0人という扱いとなっている。

13日の図-2では、ナゴヤドームを含むメッシュにピンポイントに集中している。これは、試合時間前にナゴヤドーム周辺に滞在していた試合の観客が入場し、試合を観戦していることを表していると考えられる。

(2) イベント地周辺における基礎集計分析

イベント地周辺にエリアを限定し、イベントの影響をさらに細かく見ることとする。集計対象メッシュを図-3に示す。イベント会場としてはナゴヤドーム(図-3中の①)、最寄駅としてナゴヤドーム前矢田駅(名古屋市営地下鉄名城線、図-3中の④)、砂田橋駅(名古屋市営地下鉄名城線、図-3中の⑤)、大曽根駅(名古屋市営地下鉄名城線、図-3中の⑥)を含む6メッシュとした。

細かくイベントの影響を見るために、1メッシュ毎にピックアップして分析を行い、①と③についてのみ考



図-3 ナゴヤドーム周辺

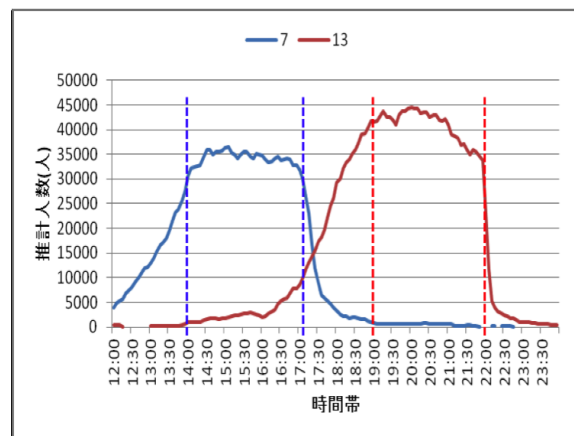


図-4 ①の時間帯別推計人数

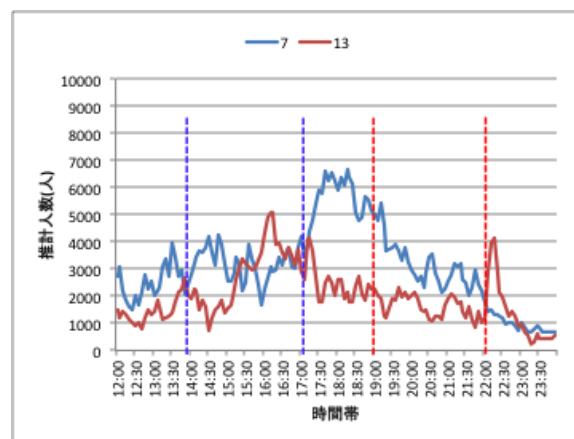


図-5 ③の時間帯別推計人数

察を行った。図-4と図-5中の青点線は7日の、赤点線は13日の試合開始時刻と終了時刻を示している。また、青線、赤線はそれぞれ7日、13日の時間帯別のメッシュ内推計人数を表したものである。ナゴヤドームが含まれているメッシュ①の時間帯別推計人数の変動を図-4に示しており、試合時刻が迫るにつれて人が徐々に集まりだしていることがわかる。試合時間中は当該メッシュ

の推計人数に大きな変動は確認できなかったが、試合終了後からは、急激に推計人数が減少している。

次に、最寄駅であるナゴヤドーム前矢田駅④に向かう最短ルート上の③に着目した。7日の場合、試合終了直後の17:30頃、推計人数が急激に増加しているが、その後暫くは推計人数にあまり変化が見られなかった。観客がナゴヤドームから移動する際に、出口の混雑により③に留まらざるをえない状況にあった可能性が伺える。また、③の周辺には大型ショッピングモールがあるため、試合終了後に観客が立ち寄った可能性も考えられる。13日の場合、試合終了時間(21:57)が大型ショッピングモールの閉店時間後であるため、観客はどこにも立ち寄ることなく帰宅したことが見てとれる。



図-6 分析対象範囲

4. 周辺人口予測モデルの構築

イベント地周辺の「混雑統計⑥」データを用いることでイベント開催前後の人が1メッシュ毎に徐々に移動していく様を確認することができた。このことを踏まえ、イベント開催地メッシュを除いたイベント地周辺に限定した周辺人口予測モデルの構築を、空間同時自己回帰モデルを用いて行う。統計ソフトである「R」を用いて行った。用いたデータは、前章で用いた、7日と13日の時間帯別メッシュ集計データを使用し、それぞれのデータでモデルを構築した。

(1) 空間同時自己回帰モデル

メッシュ内人口(人)について空間同時自己回帰モデルを用いて周辺人口予測モデルを構築する。まず、空間的自己相関を表す、空間重み付け行列 W を作成した⁸⁾。メッシュの中心点の緯度と経度を用いて、2つ隣のメッシュまで影響があると仮定し、行列を作成した。あるメッシュ内の人口は、空間同時自己回帰モデルを用いて以下のように表すことができる。式(1)中の ρ と X は、それぞれ空間ラグのパラメータと説明変数である。

$$y = \rho W y + X \beta + \varepsilon \quad (1)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

また、未知パラメータ β は最尤法にてそれぞれ推定を行った。

(2) 分析範囲

基礎集計分析において着目したメッシュ以外にも、イベントの影響が確認できるメッシュが存在している可能性を考慮し、イベント地を中心とした約2.5km四方の範囲を分析範囲とした。その範囲を図-6に示す。な

お、本研究はイベント地周辺の流動人口を予測するため、開催地メッシュは分析に用いていない。

(3) 空間同時自己回帰モデルの説明変数

使用した説明変数の詳細を表-2に示す。イベント後ダミーは、説明変数の「時間差」においてイベント前後を示したものである。平常時の推計人数については、今回の用いたデータには平常時のデータはないため、7月21日のデータ(別途、取得済み)を平常時のデータとして利用した。

(4) 空間同時自己回帰モデルの結果と考察

7日と13日のデータを用いて構築した空間同時自己回帰モデルの各パラメータを表-3に示す。7日のデータを用いた分析結果では、定数項を始めとするすべての推定値が統計的に有意となった。また、AICは116,720であり、重回帰分析を行った場合のAICよりも下がっており、重回帰モデルよりも当てはまりが良いモデルであるといえる。説明変数については、「時間差」や「イベント地距離」、「最寄駅距離」の推定値は負となると予測されたが、推定値は「時間差」のみ正となった。この場合、イベント開催時間から離れば離れるほど、メッシュ内人口が多くなるという傾向にある。これは、7日が休日であり、イベントが昼間に開催されたために、ナゴヤドーム周辺にある大型ショッピングモールへ夕方にかけて人々が集まってきたためと考えられる。

13日のデータを用いて分析を行った結果は、「開催後ダミー」と「道路面積」の推定値が統計的に有意にはならなかった。これは、13日が平日であり、イベントがナイターゲームであったためと考えられる。平日ということもあり、公共交通機関での移動が多いことや、イベント終了後にすぐ帰宅してしまったことが、この

表-2 説明変数の概要

説明変数	単位	概要
開催後ダミー	-	時間帯がイベント後であれば1, そうでなければ0
時間差	分	イベント開始時刻までの, または, イベント終了時刻からの時間差の絶対値
イベント地距離	m	当該メッシュからイベント地までの最短直線距離
最寄駅距離	m	当該メッシュから最寄駅までの最短直線距離
イベント規模	万人	開催イベントの(予定)観客数
平常時推計人数	人	上記の仮定に従って変数を作成
道路面積	km ²	当該メッシュ内の主要道路面積

表-3 空間同時自己回帰モデル推定結果

	7 日		13 日	
	推定値	Z 値	推定値	Z 値
定数項	237.93	4.94	351.98	10.83
開催後ダミー	130.76	6.15	21.98	1.35
時間差	0.72	3.02	-0.52	-2.93
イベント地距離	-0.14	-5.71	-0.21	-12.27
最寄駅距離	-0.18	-5.70	-0.14	-6.78
平常時推計人数	0.76	64.96	0.86	92.15
道路面積	-737.85	-6.34	83.36	-0.98
ρ	0.31	11.27	0.27	12.54
AIC	116,720		125,540	
AIC(重回帰)	116,930		125,730	
サンプル数	7,265		8,090	

ような結果となった原因であると考えられる。

また、図-7と図-8は、それぞれのモデルの推計値を用いて算出したメッシュ内人口(以降、予測値)と実際のデータにあった推計人数(以降、実測値)をプロットしたものである。モデルの評価をする際に、AICとは別にRMSE(Root Mean Square Error)も用いた。RMSEとは、平均二乗誤差のことであり、計算式は式(2)の通りである。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - x_i)^2} \quad (2)$$

真値として実測値を用いてRMSEを算出した。一般的に相関の高いモデルはRMSEが小さくなると言われている。今回構築した2つのモデルのうち、相関の高いものは、13日のデータを用いて構築したモデルであった。いずれにおいても、実測値が予測値を上回る結果となった。

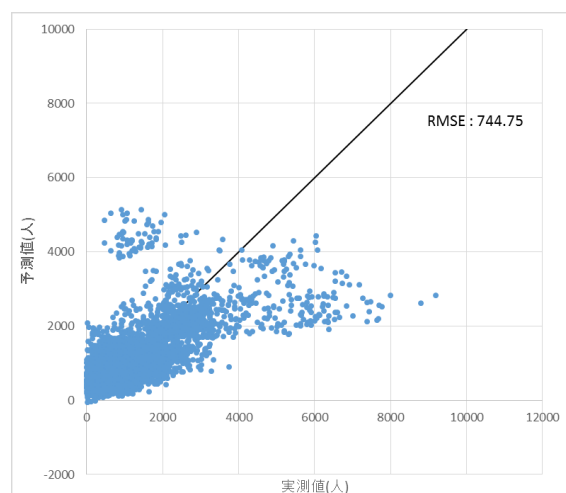


図-7 実測値と予測値の相関(7日)

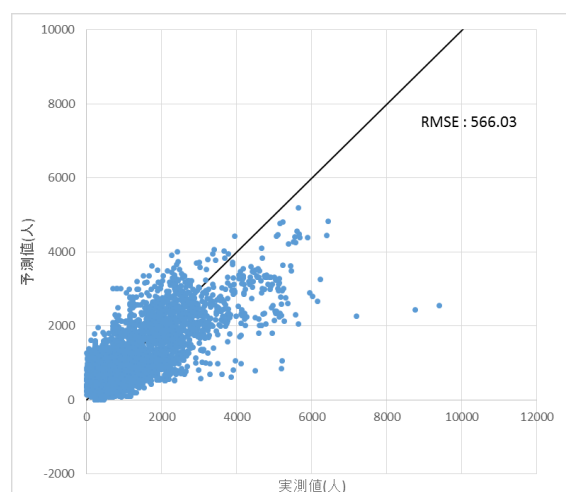


図-8 実測値と予測値の相関(13日)

45度線から大きく外れたデータ(外れ値)にはどのような性質があるのかを検証するためにGIS上にて実測値と予測値との差について可視化を行った。なお、時間帯は、イベント開催時間の2時間前と2時間後、イベ

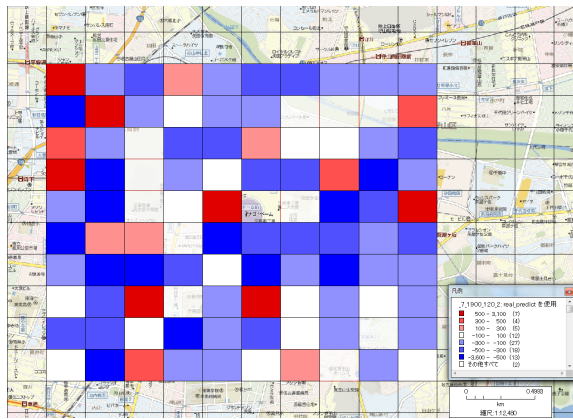


図-9 実測値と予測値の比較 (7日, イベント開始2時間後)

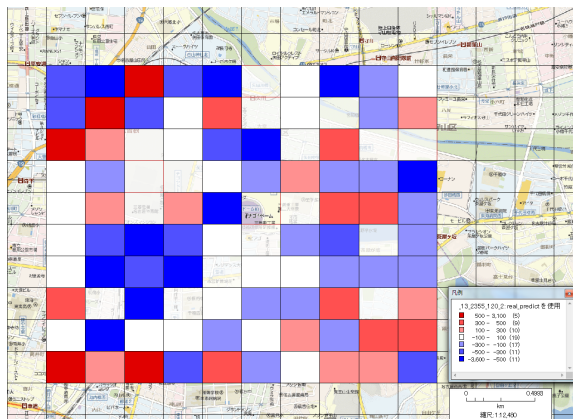


図-10 実測値と予測値の比較 (13日, イベント開始2時間後)

ント開催時間中について可視化を行った。実測値の方が大きければ赤色が、予測値のほうが大きくなれば青色が濃くなる様に可視化を行った。今回は、イベント終了から2時間後のデータについてのみ考察する。まず、7日のモデルについては、全体的に予測値のほうが大きく算出される傾向にあった。特にイベント終了2時間後である図-9において顕著に現れている。これは、おそらく正の推計値となっている「時間差」や「開催後ダミー」が大きく影響しているものと考えられる。13日のデータを使用して構築したモデルの予測結果は、全体的に白色のメッシュ、つまり予測値と実測値の差異が少ないものが多くなった。7日と異なり、「時間差」の推定値が負となったことが原因であると考えられる。

5. 結論

(1) 得られた知見

本研究で利用した「混雑統計®」データは、メッシュ毎に集計されたデータであるため、個人の移動を細かく追っていくことは不可能なデータである。全体で見た場合は、他のイベントが開催されていないメッシュ内

人口の変動に相殺されてしまい、イベントの影響が確認できない。そこで、イベント地周辺に絞って分析を行った結果、イベントの影響を確認することができた。イベント地周辺について1メッシュずつ集計を行った場合には、イベント開始時刻前までに徐々にイベント地に人々が集まりだし、イベント終了後からはイベント開催時間やイベント開催範囲によって異なる人の動きがあった。また、メッシュ内人口が波及していくことが確認できた。

それを踏まえた上で、利用可能性の一つとして空間同時自己回帰モデルを用いたイベント地周辺の人口予測モデルの構築を行った。この際に、分析対象範囲をイベント開催地周辺の約2.5km四方の範囲とし、空間重み付け行列を用いて空間同時自己回帰モデルを構築することで、重回帰モデルよりもモデルの当てはまり度合いを示すAICが下がることが確認できた。そして、大まかではあるが、実際のメッシュ内人口と同じような分布をすることが確認できた。人々の位置情報が記録されたビッグデータに対しこのモデルを用いることで、イベント開催地・影響範囲が限定されるが、新規にイベントを開催した場合においても容易な計算で混雑地域のメッシュをある程度判別でき、警備員の配置や地下鉄の増便、交通規制などの対策を講じるための基礎的資料として用いることが可能になると考えられる。

(2) 今後の課題

本研究では、「混雑統計®」データの時間帯別メッシュ内人口データを用いて、イベント地や最寄駅との位置関係や時間帯、メッシュ内の道路面積に着目してモデル構築を行った。より高い精度をもつモデルや様々なイベントに対応のできるモデルを構築するためには、分析対象範囲中のメッシュ内における交通の供給側に着目した変数や地理的な情報を表した変数を新たに導入することや、分析対象範囲の再検討、分析を行うデータ量を増やすことが必要である。

謝辞： 本研究は、文部科学省科学研究費 基盤研究(B)[研究課題番号 25289161]の支援により実施したものである。

参考文献

- 1) 明石市民夏祭り事故調査委員会: 第32回明石市民夏祭りにおける花火大会事故調査報告書, 明石市, 2013.
- 2) 総務省: 情報通信統計データベース, <http://www.soumu.go.jp/johotsusintokei/index.html>, 2013.11.06.
- 3) 株式会社ゼンリンデータコム: 混雑度マップ, <http://lab.its-mo.com/densitymap/>, 2014.01.23.
- 4) 羽藤英二・朝倉康夫・喜村裕二: 移動体システムを用いた大規模イベント時の交通行動分析, 土木計画学研究・講演集, Vol.22, No.1, pp. 409-412, 1999.

- 5) Hayano, R. and Adachi, R.: Estimation of the total population moving into and out of the 20 km evacuation zone during the Fukushima NPP accident as calculated using "Auto-GPS" mobile phone data, Proceedings of the Japan Academy Series B, Vol. 89, No. 5, pp. 196-199, 2013.
- 6) プロ野球 Freak: 中日ドラゴンズの観客動員数,
<http://baseball-freak.com/audience/12/dragons.html>,
(2013.10.15)
- 7) 株式会社ナゴヤドーム:
ナゴヤドームオフィシャルホームページ,
<http://www.nagoya-dome.co.jp/>, (2013.10.15)
- 8) 古谷知之: R による空間データの統計分析, 朝倉書店,
2011

(2014. 08. 01 受付)

A Study of Spatial Auto-regression Model for Estimation of Migrant Population
around Event Spot Using "Konzatsu Tokei" Data

Takumi BAN, Tomotaka USUI, Toshiyuki YAMAMOTO, Takuyuki MORIKAWA