データから都市を診断して将来を考える

嚴　先鏞

近年都市の問題を解決するためにデータの活用の重要性が増えつつある．都市計画分野のみならず，多くの分野の研究者にとってこのようなデータを賢く活用することは重要な課題である．本稿では，さまざまなデータから都市を診断して問題を解決することを目指して行ってきた研究の事例を紹介しながら，どのようにデータと付き合えばよいかについて 1) データがもつ情報を最大限に活用する，2) さまざまなデータの融合から発見する，3) 現状だけでなく将来を考えるという三つの観点から意見を述べる．

キーワード：データ分析，都市計画，都市解析

1. はじめに

筆者はさまざまなデータと都市解析手法に基づいて，われわれが住んでいる都市を診断して，問題点の発見と解決に貢献するという大きい目標をもって現在研究を行っている．その際に不可欠となるのがさまざまなデータの活用である．近年，ビッグデータは常識となっており，ミクロレベルでの詳細なデータが増えつつある．たとえば，センサーから収集された人の流れに関するデータ，ミクロレベルの土地利用データ，新聞やソーシャルネットワーキングサービス (SNS) からのテキストデータなどが挙げられる．

研究者としてこのようなデータを賢く活用することは重要な課題である．新しい分析手法を学んで適用することは大事であるが，その前にデータとどのように付き合えば良いかについても考えてみる必要がある．そこで，本稿では，筆者がデータに基づき都市に関する研究をしながら感じたデータとの付き合い方についての経験を共有したい．具体的には，さまざまなデータを用いる中で重要と感じた以下の三つのポイントについて，これまでに行った研究事例に基づき紹介する．

- データがもつ情報を最大限に活用する．
- さまざまなデータの融合から発見する．
- 現状だけでなく将来像を考える．

2. データがもつ情報を最大限に活用する

近年のデータは膨大な情報をもっていることが多く，元データがもつ情報を最大限に活用することが重要である．しかし，実際の分析において元データがもっていた豊かな情報が失われてしまうことが多くある．そこで，筆者がデータのもつ情報をより活用することに力を入れた，用途間の空間関係を用いた土地利用混合度指標に関する研究 [1, 2] と複数地点の歩行者量データに基づいた歩行者分布の把握に関する研究 [3] を紹介する．

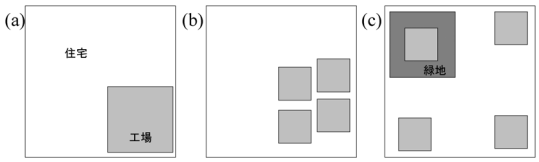


図 1　異なる土地利用混合形態の例

2.1 用途間の空間関係を活用した土地利用混合度

まず，博士論文の事例を少し紹介したい．博士論文のタイトルは「用途間の相互関係・混合度に着目した土地利用構造の定量化と評価に関する研究」である．都市空間にある住宅，商業，工業，公園などからなる土地利用用途がどのように混合されているのかが研究の出発点である．図 1 のように住宅と工場が分布している単純な地区を考える．(a) のように工場が住宅から完全に分離されているパターンに比べて (b) のように工場用途が細かく分離されているパターンでは工場による騒音や臭いなどの悪影響がより発生する可能性が高い．(c) のように第三の用途により囲まれて，バッファの役割を担っている場合，隣接による影響はないものの，騒音や臭いなど距離逓減の少ない事象が影響する可能性もある．このような土地利用の混合状況は量が同じであっても空間的な分布によって実際の都市空間での環境は異なる．最近は図 2 のように敷地単位での土地利用データを入手することが可能となり，豊かな空間データが整備されている．しかし，空間的な分布は考慮されず，量だけで定量化されることが多く，

おむ　そんよん
筑波大学システム情報系
〒 305–8573 茨城県つくば市天王台 1–1–1
eom.sunyong.fw@u.tsukuba.ac.jp

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

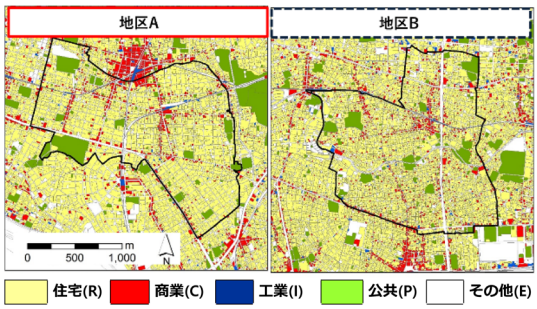


図 2　敷地単位の土地利用データの例

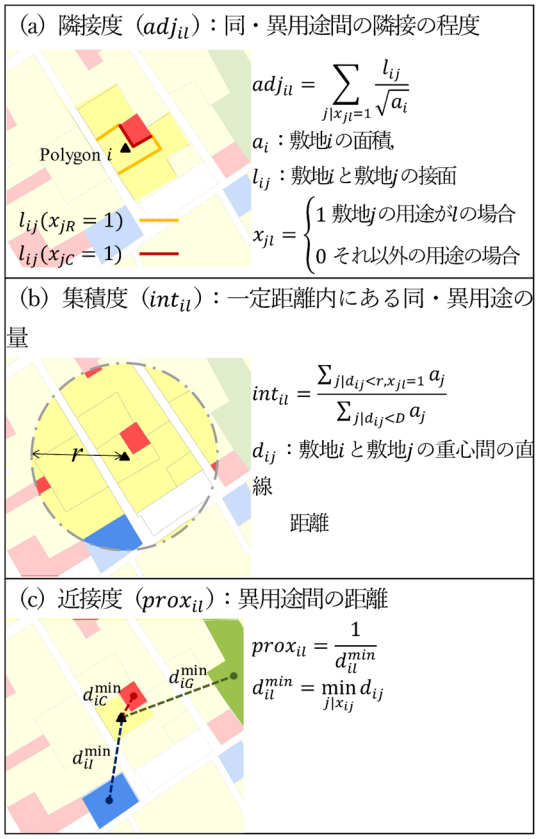


図 3　土地利用パターンの例

その代表例として用途間の量のバランスを測るエントロピー指標 [4] が挙げられる．

ここで元の土地利用データがもっている空間情報を最大限に活用した指標を開発したのが博士論文の最初の内容である．図 3 のように敷地 i における用途 l に対する混合度を，隣接度，集積度，近接度の三つの空間関係から定義する．その後，一定の範囲をもつ地区（たとえば町丁目）単位である 2 種類の用途間 (k, l) の混合度である隣接性 (Adj_{kl})，集積性 (Int_{kl})，近接性

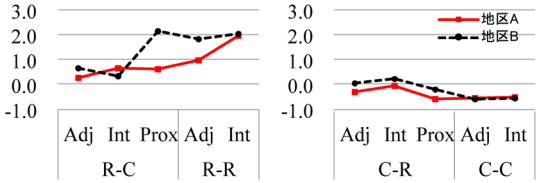


図 4　混合度指標による地区 A と B の比較

図 5　豊田市の歩行者量観測地点

$(Prox_{kl})$ を式 (1) ～ (3) で定義する．

$$Adj_{kl}=\frac{\sum_{i|x_{ik}=1}a_i adj_{il}}{\sum_i a_i} \tag{1}$$

$$Int_{kl}=\frac{\sum_{i|x_{ik}=1}a_i int_{il}}{\sum_i a_i} \tag{2}$$

$$Prox_{kl}=\frac{\sum_{i|x_{ik}=1}a_i prox_{il}}{\sum_i a_i} \tag{3}$$

これらの指標を用いることにより，用途別の構成比はほぼ同じである図 2 の二つの地区の混合度の違いが図 4 のように把握できる．すなわち，元の敷地単位のデータがもっている情報を定量化して最大限に活用した一つの例であり，博士課程の研究生活から学んだ重要なデータとの付き合い方である．

2.2　量・空間・時間を考慮した歩行者分布指標

次は，博士課程での経験を生かしてデータがもっている情報を最大限に活用した最近の研究を紹介する．近年都市空間に設置されたセンサーからさまざまな情報が収集されている．その中で歩行者の活動に関する情報は，まちのにぎわい，健康促進，安全性などの観点から活用されている．カメラによって複数地点・時点で観測されている歩行者通行量データもその一つである．しかし，量，時間，空間の属性をもっているにもかかわらず，実際の現場では量のみ用いられることが多い．そこで，単純な量のみならず，歩行者の空間・時間的な分布まで把握できる指標を開発した [3]．

対象地域である豊田市では図 5 のように 2008 年から中心市街地を通過した歩行者の人数をカウントするためのカメラを設置して運営している．図 5 の 20 箇

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

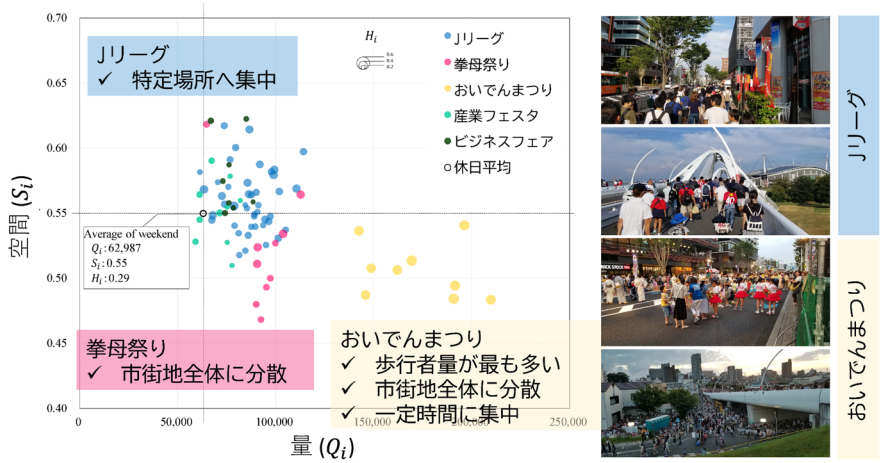


図 6　歩行者分布指標によるイベント別の特徴の比較

所の中で，新設された，または測定の中止期間があった 7 箇所を除き，2008 年からのデータの蓄積のある 13 箇所（青い地点）のデータを用いた．収集されたデータは 1 時間単位にそれぞれのカメラで確認した通過人数である．

そこで，記録された場所と時間の情報を活用した歩行者分布指標の開発を目的として研究を行った．当時はデータ分析の専門家でなくても使える手法の開発も一つの重要な要素であったため，空間的・時間的な偏りをジニ係数によって定量化することで，容易に分析できる指標を提案した．量指標 (Q_i) は i 日における歩行者カウントの総和，空間指標 (S_i) は複数の測定地点における分布の不均等性，時間指標 (H_i) は複数の測定時間における分布の不均等性を表す．

$$Q_i = \frac{1}{n}\sum_k c_{ik} \tag{4}$$

$$S_i = \frac{1}{2n^2 Q_i}\sum_k\sum_l |c_{ik} - c_{il}| \tag{5}$$

$$H_i = \frac{1}{2m^2 Q_i}\sum_t\sum_s |v_{it} - v_{is}| \tag{6}$$

ここで，c_{ik} は i 日の地点 k における歩行者カウント数，v_{it} は i 日の時間帯 t における歩行者カウント数，n, m はそれぞれ測定地点数，測定時間帯数である．

図 6 は，豊田市で行われたこれまでのイベントを対象に提案した指標を適用した結果である．第一に，イベントのある日は，休日の平均と比べてすべての指標が大きく，歩行者がイベント時間帯と場所に集中していることがわかる．第二に，大規模な祭りの日は，休日より S_i が小さい傾向があり，都心全体に歩行者が広がっている傾向がある．第三に，Q_i からは，拳母祭りと J リーグの差はあまりなかったものの，S_i では，J リーグの方が大きい値を示している．拳母祭りは都心全体において来訪者が多い一方，J リーグは歩行者が一部の地点のみ偏ることがわかる．量のみならず，元データがもっている空間的・時間的情報を活用することで，歩行者の分布を把握し，イベントの効果の評価，イベントと連携した中心市街地活性化の促進に寄与できると考えられる．

3.　さまざまなデータの融合から発見する

近年活用されているビッグデータは情報量が非常に多いが，さまざまな理由で一つのデータから必要なすべての情報が得られない場合も多い．たとえば，携帯位置情報に基づいた滞在人口はプライバシーの問題で得られる情報には限界があり，SNS から得られる情報も不完全な情報であることが多い．データを用いる研究者としては使用データがもっている限界を明確に理解することは当然であるが，その限界を克服する方法を探すことも重要である．そこで，筆者が試みたのが複数のデータの融合である．その例として携帯位置情報からの滞在人口データと前に紹介した歩行者量観測データを融合した研究の事例を紹介する [5]．この研究は，複数の交通ビッグデータを用いた分析により一つのデータだけでは捉えられない変化を把握することを目的としている．

滞在人口データと歩行者量観測データは人出の変化を把握する際に有用であるが，それぞれ限界がある．まず，COVID-19 でその活用が急速に広がっている位置情報に基づいた滞在人口データは，携帯の基地局の

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

図 7　12 時帯の滞在人口と歩行者量データの比較

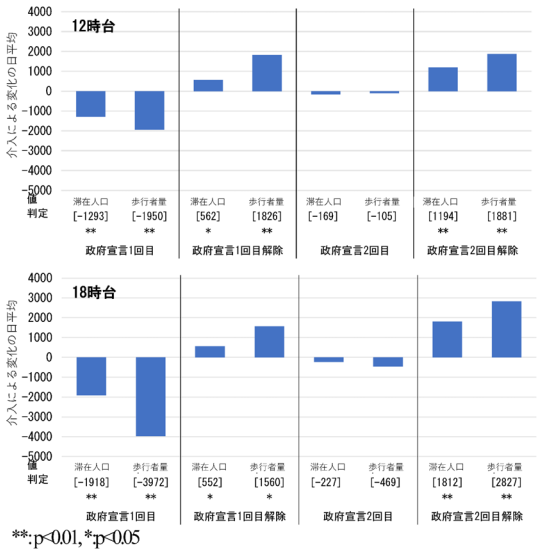


図 8　BSTS による宣言開始と解除の効果の比較

位置情報に基づいて推定される [6]．しかし，プライバシーの問題により滞在していた人々に関して一部の情報は把握できない．滞在した人数と滞在時間が同時に考慮して推定されるため，10 人が 10 分間滞在した場合，1 人が 100 分間滞在した場合の滞在人口は同じく 100 人・分として表示される．一方，歩行者量観測データは，通過した人は滞在時間に関係せず，1 人として記録される．したがって同じ人数が観測されても滞在時間により市街地での人出の状況は異なる可能性もある．図 7 は，豊田市の中心市街地の 12 時帯において 2019 年と 2020 年の 7 日移動平均を求めたうえで，その差を表したものである．すなわち負の値は人出が減少したことを意味する．1 回目の緊急事態宣言の解除後を見ると，歩行者量観測データはすぐに回復しているものの，滞在人口データは宣言前の水準を下回っていることが見られる．

二つのデータにおける人出の増減と感染拡大防止対策の関係を調べるため，ベイズ構造時系列モデル (BSTS) [7] を用いる．このモデルを用いることで，介入の効果を定量的に把握することが可能である．分析には R の CausalImpact package を使用した．図 8 は緊急事態宣言による人出変化の分析結果の一部を示している．1 回目の緊急事態宣言により，滞在人口データと歩行者量データはともに有意な減少が確認された．12 時帯の結果を見ると宣言の解除により，人出は増加したが，その増加量において二つのデータに差がみられる．滞在人口データは減少した量に比べて増加した量は小さい．一方，歩行者量データは減少した量と増加した量がほぼ同じである．まちなかを歩く人の量（歩行者量データ）の回復幅よりも，滞在時間を加味した滞留人口の回復幅が小さいということは，滞在時間が短くなっている可能性があることを示唆している．一方，18 時の場合には歩行者量データの増加も減少より小さいことが読み取れる．すなわち，人の量も宣言前の水準を下回っていることがわかる．

これは 2 種類のデータを比較することでしか把握できない貴重な知見である．年々増加しているデータを用いて現実の問題を発見する際に，この研究はデータの限界の理解と複数のデータの融合の重要性を示すことができたと思われる．

4.　現状だけでなく将来像を考える

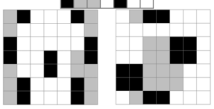
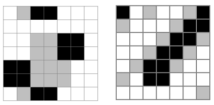
都市計画の研究者として，どのような都市がより暮らしやすいかの質問に答えることは重要だと思う．データから都市の実態や現実の問題を発見できるが，これからの都市のデータはあるわけでなく，これまでのトレンドが続くとも言い切れない．将来を考える一つの方法として最適化手法が有用である．それと関連した研究として用途間の相互作用を考慮した土地利用配置モデル [2] と感染症における隔離施設の配置モデル [8] の研究成果を紹介したい．

4.1　用途間の相互作用を考慮した土地利用配置モデル

2 節で紹介した隣接性，集積性，近接性の土地利用混合度指標を用いて現状の把握は可能である．混合の状況により，実際の住環境に正・負の影響が発生する．したがって，さまざまな用途間の隣接性，集積性，近接性において評価基準が与えられたとき，意思決定者における重要な課題は，「どのように土地利用を配置すれば負の関係をもつ混合を防ぎ，正の関係をもつ混合を増やすことができるか」であるといえる．そこでこの研究では，用途間 k と l の混合度（Adj_{kl}, Int_{kl}, $Prox_{kl}$）における評価基準（w_{kl}^{Adj}, w_{kl}^{Int}, w_{kl}^{Prox}）が与えられたとき，その評価を高くする土地利用パターン

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

表 1　混合度に対する評価別の最適土地利用パターンの特徴

	混合度に対する評価 w_{WB}^{Adj}	w_{WB}^{Int}	w_{WB}^{Prox}	最適パターンの特徴	パターン例
W1	+	+	+	・用途 B（黒のセル）が用途 W（白のセル）の中で分散して均等に分布しているパターン	
W2	-	+	+	・用途 B が小さい塊を形成しながら均等に配置されることにより，用途 W と B の間の隣接を防ぎながら集積性と近接性を大きくするパターン	
W3	-	+	-	・用途 B が大きい塊を形成して均等に配置されて用途 W と B の間の隣接と近接を防ぎながら集積性を担保する ・用途 W と B 両方とも負の関係をもたない用途（灰色のセル）がある場合，バッファ空間として配置するパターン	

の特徴を土地利用配置モデルに基づいて明らかにした．

評価基準が正であれば，その混合度が大きければ大きいほど土地利用パターンの評価値が高く，負であれば，その混合度が小さければ小さいほど土地利用パターンの評価値が低いと考えると，以下のような多目的問題として考えることができる．

$$\text{Max } w_{kl}^{Adj} Adj_{kl} \tag{7}$$

$$\text{Max } w_{kl}^{Int} Int_{kl} \tag{8}$$

$$\text{Max } w_{kl}^{Prox} Prox_{kl} \tag{9}$$

多目的問題を解くために Reference point 手法を用いている [9]．Reference point 手法の特徴は，得られた解はパレート最適であるということであり，以下の式によって定式化できる．

$$\text{Min} \sum_r w_r \left[\frac{z_r^* - z_r}{z_r^* - g_r} \right]^\gamma \tag{10}$$

ここで，$z_1, z_2, z_3, \cdots, z_M$ は個別目的関数であり，z_r^* は z_r の理想の値（たとえば z_r の最大値），g_r は目的関数 r における Reference point（目標値），w_r は目的関数 r の相対的な重要度である．γ は目的関数の違反が大きくなることに対するペナルティーであり，ここでは $\gamma = 4$ と設定した．

三つの用途で構成されている仮想のグリッド空間に土地利用配置モデルを適用して混合度に対する評価基準を満たす規範的な土地利用パターンの特徴を明らかにした．土地利用パターンを求める際には，ランダム交換によるヒューリスティックプロセスを用いた．詳細については論文 [2] を参照にしていただきたい．

表 1 は混合度に対する評価による最適パターンの特徴の例である．たとえば，コンビニが近くにあるのは良いが隣接していると騒音の問題があるため，ある程度離れてほしいという W2 のように隣接性には負の評価，近接性と集積性には正の評価になる．それに対応する最適土地利用パターンの特徴は，用途 B が細かい塊を形成して均等に分布していることである．これらの結果を用いることで，これからの新しい地区計画におけるガイドラインの作成や，変化しつつある混合に対する評価基準に適した土地利用のパターンの特徴の理解が可能である．

4.2　感染症における隔離施設の配置モデル

COVID-19 のワクチンの普及にかかわらず人流の抑制や感染者の隔離などの対策は感染拡大防止のために重要である．隔離施設を活用した戦略は，韓国，中国，シンガポールなどで有効な効果を示した．韓国では研修施設，大学宿舎，ホテルなど公共と民間施設が生活治療センターとして利用されているが，利用可能な施設の確保とスケジューリングの問題で，十分な数の施設の確保が困難であった．したがって，この研究では，感染症に効率的に対応するための地域の施設配置・活用モデルを提案した．具体的には，韓国を対象に利用可能な施設の調査，感染症状況のシナリオの作成，施設選択・スケジュールモデルを提案した．ここではモデルとその結果について簡単に紹介する．

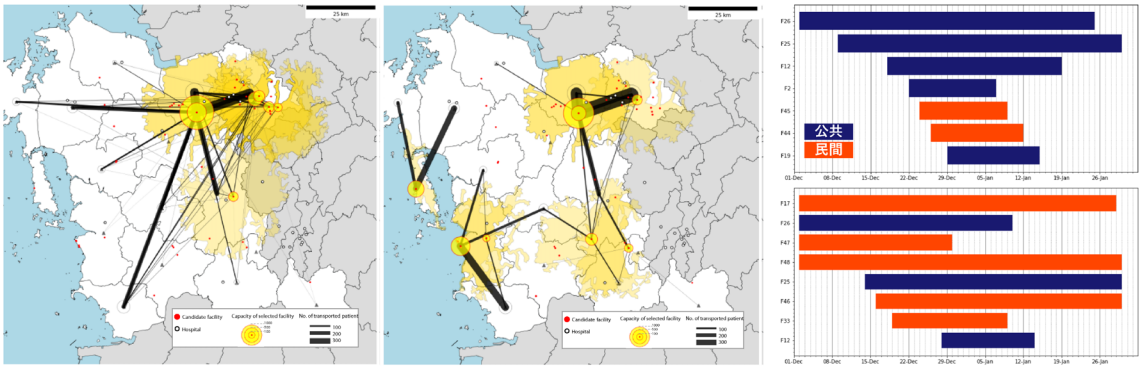
提案したモデルは，利用可能な施設の分布と感染者数の推定結果が与えられた場合にどの施設を何時から何時まで隔離施設として利用すれば良いかに答える問題であり，以下のように定式化できる．

$$\underset{\text{Subject to}}{\text{Minimize}} \sum_{i,t} (x_{it}(R_i C_i + S_i) + F y_{it}) + \sum_{i,j,t} p_{ijt} V_{ij} \tag{11}$$

$$\sum_i p_{ijt} = n_{jt}, \forall j, t \tag{12}$$

$$\sum_j p_{ijt} - M x_{it} \leq 0, \forall j, t \tag{13}$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.



(a)集中型の施設選択　(b)分散型の施設選択　(c) スケジュール（上：集中，下：分散）

図 9　隔離施設配置・スケジュールの結果の例

$$x_{it} - x_{i,t-1} \leq y_{it}, \forall i, \forall t | t > 0 \tag{14}$$

$$x_{i0} = 0, \forall i \tag{15}$$

$$\sum_{s=max(0,t-14)}^{t} p_{ijs} - C_i X_{it} \leq 0, \forall i, j, t \tag{16}$$

$$\sum_{t} y_{it} \leq 1, \forall i \tag{17}$$

ここで，

x_{it}：施設 i を t 日に運営すれば 1，さもなくば 0

y_{it}：施設 i を t 日に転換すれば 1，さもなくば 0

p_{ijt}：地域 j から施設 i まで t 日に搬送する感染者数

n_{jt}：地域 j の t 日の感染者数の推定結果

R_i：施設 i を隔離施設に転換する際の難易度

C_i：施設 i の病室数

S_i：施設 i の運営に必要なスタッフの人数

V_{ij}：地域 j から施設 i まで搬送するためのコスト

F：施設の転換コスト

M：十分に大きい定数

である．式 (11) の目的関数は，施設の運営コストと感染者の搬送コストの総和の最小化である．式 (12) は発生したすべての感染者を施設に隔離することを，式 (13) は 1 名以上が隔離されると隔離施設として指定することを意味する．式 (14) は施設を隔離施設に転換する時点を表し，式 (15) では $t = 0$ のときに隔離対象者がいないことを仮定する．式 (16) は，隔離期間 15 日を考慮したうえで搬送された感染者数が施設の病床数を超えないこと，式 (17) は隔離施設から解除されたら再度指定しないことを意味する．

施設の所有者と用途により，R_i を設定した（公共・研修施設：1.0，公共・宿舎：1.5，民間・研修/宿舎：1.5，民間・ホテル：2.0）．スタッフについては，病床数により 100 室未満は 50 名，100 室～199 室は 65 名，200 室以上は 80 名と設定した．最後に搬送コストは 2 種類を設定し，地域間の移動コストを考慮しないケース（集中型）とできるだけ隣接の自治体内で搬送するケース（分散型）に分けて結果を考察した．

図 9 は，感染症が実際より拡大した場合を想定したシナリオの結果を示している．集中型は対象地域の北西部に集中して施設が指定される一方，分散型は複数の拠点をもって施設が指定されていることがわかる．分散型の場合，公共施設がない地域における民間施設の利用が多く指定され，全体で確保される病床数も隔離対象者数より多くなる．搬送コストを減らすために施設の運営費が大きくなるが，搬送を含めた医療体制を考えると距離の短縮効果は重要であると考えられる．

5. おわりに

本稿では，データから都市を診断して将来を考えるために行った研究の事例を紹介しながら，データとの付き合い方について筆者の意見を述べた．博士論文の審査のとき，「いろいろな分析を行っていることはわかるが，あなたの哲学は何でしょうか」という質問を受けた．今考えると当時の私の発表では，分析した結果を述べていただけで，これからの研究者としての夢や覚悟のようなものが見えなかったようだ．本稿で述べたのは筆者のデータ分析に対する哲学ともいえる．データを活用するためには高度な分析手法も必要となるが，そもそもデータと付き合うための心がけとして参考にしていただきたい．

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

参考文献

[1] 嚴先鏞，鈴木勉，“用途間の隣接性・集積性・近接性の観点から見た混合土地利用パターンの定量化，”都市計画論文集，**51**, pp. 867–874, 2016.

[2] S. Eom, T. Suzuki and M.-H. Lee, “A land-use mix allocation model considering adjacency, intensity, and proximity,” *International Journal of Geographical Information Science*, **34**, pp. 899–923, 2020.

[3] S. Eom and Y. Nishihori, “How weather and special events affect pedestrian activities: Volume, space, and time,” *International Journal of Sustainable Transportation*, **16**, pp. 462–475, 2022.

[4] K. Kockelman, “Travel behavior as function of accessibility, land use mixing, and land use balance: Evidence from San Francisco Bay area,” *Transportation Research Record: Journal of the Transportation Research Board*, **1607,** pp. 116–125, 1997.

[5] 西堀泰英, 嚴先鏞, 佐々木邦明, 加藤秀樹, “交通ビッグデータを用いた地方都市中心市街地の人出等に対する COVID-19 感染拡大防止対策の影響分析，”都市計画論文集，**56**, pp. 834–841, 2021.

[6] M. Terada, T. Nagata and M. Kobayashi, “Population estimation technology for mobile spatial statistics,” *Docomo Technical Journal*, **14**, pp. 16–23, 2013.

[7] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy and S. L. Scott, “Inferring causal impact using bayesian structural time-series models,” *Annals of Applied Statistics*, **9**, pp. 247–274, 2015.

[8] S. Eom and E. Byun, “Local resource allocation model for quarantine facility in response to infectious Diseases,” In *Research Abstracts on Spatial Information Science CSIS DAYS 2021*, D09, 2021.

[9] T. J. Stewart and R. Janssen, “A multiobjective GIS-based land use planning algorithm,” *Computers, Environment and Urban Systems*, **46**, pp. 25–34, 2014.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.