

方向統計学を用いた道路ネットワークの分析

長崎 滉大¹・中西 航²・朝倉 康夫³

¹学生非会員 東京工業大学 工学部 土木・環境工学科 (〒 152-8552 目黒区大岡山 2-12-1)
E-mail: k.nagasaki@plan.cv.titech.ac.jp

²正会員 東京工業大学助教 環境・社会理工学院 土木・環境工学系 (〒 152-8552 目黒区大岡山 2-12-1)
E-mail: nakanishi@plan.cv.titech.ac.jp

³正会員 東京工業大学教授 環境・社会理工学院 土木・環境工学系 (〒 152-8552 目黒区大岡山 2-12-1)
E-mail: asakura@plan.cv.titech.ac.jp

道路網を構成するリンクやノードの接続関係を用いて道路網の形態を分析する研究は行われてきたが、特定の都市施設周辺の道路網に焦点を当て、各道路の実延長や偏角などから道路網の評価を行う研究はほとんど存在しない。本研究では、角度データを扱う方向統計学を用いた新たな道路網の評価手法を構築した。まず、駅周辺の各道路の偏角に着目したグラフを作成した。次に、偏角の確率分布のパラメータを推定する方法を開発するとともに、推定パラメータを用いて駅周辺空間の道路網をクラスタリングする方法を提案した。東京圏一都三県の駅に適用し、一連の手法の一定程度の有用性を示すとともに、問題点や課題を示した。

Key Words : *directional statistics, urban morphorogy, road networks, classify, estimating distribution*

1. はじめに

道路網に関する研究について、各道路の繋がりや延長、偏角などを用いたネットワークの形の評価に関するものは少ない。時系列別でロンドン市内の道路の総延長や交差点の数に焦点をあてて道路網の成長の限界を予測する研究¹や、世界100の都市を対象に、個々の道路の偏角について分析した研究²がある。後者の研究では、角度データを扱う方向統計学のヒストグラムであるバラ図（詳細は後述）を用いて解析を行っている。この研究では偏角の指標としてバラ図のみから得られる指標を使っており、分布の推定を行っていないため、同じバラ図が得られても性質としては全く違う道路網が同じ評価値を持つという問題がある。また、都市全体を対象にしている研究がほとんどであり、駅周辺のような特定の都市施設周辺の道路網の解析はほぼ行われていない。さらに、道路網が都市の発展にどういった影響を与えているのかという点に展開した研究もほとんど行われていない。

これらを踏まえて、本研究では方向統計学の手法を用いて、駅周辺道路網を構成する各道路区間の向き（偏角）に着目し、偏角の確率分布を推定するとともに高次元の解析を行って、道路網全体の形状を評価できる手法を考案する。さらにその手法を東京都市圏の一都三県の各駅

周辺の道路網に適用し、その類型化を行う。このことにより方向統計学に基づく新しい手法で駅周辺道路網の解析と類型化を行うことを目的とする。本論文では、偏角のグラフの作成、分布の推定、分布のパラメータを用いたクラスタリングの手法の説明を行う。

2. 分析手法

(1) データ

道路データは国土地理院が発行している"数値地図(国土基本情報)"のデータであり、一都三県に存在する駅が含まれる157個の2次メッシュについて、2018年11月26日時点で購入可能であった最新のデータである。駅データは国土交通省が発行している"国土数値情報「鉄道時系列」平成29年N05-17 GML"に含まれるデータを使用した。このデータにおいて駅は一つの点として記述されている。駅データは路線別に存在するため、データ上で同じ駅名でかつ同じ駅前空間を共有していると考えられる駅は複数の座標の重心を取り一つの駅とした。この処理により駅数は1561となった。

これらのデータを用いて、駅を中心として、半径1000m以内に道路（リンク）の少なくとも一つの端点が

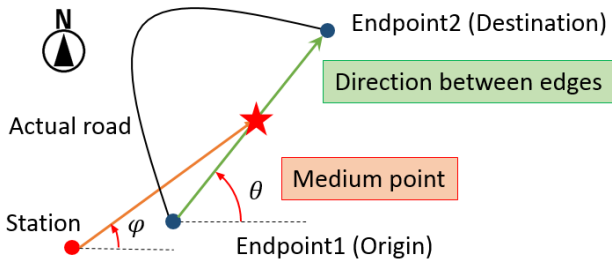


図-1 道路データの処理

存在する道路の端点の座標，実延長を抽出した．端点の定義は交差点である．

以後，この二つの交差点の座標で記述される道路を"道路リンク"と呼ぶ．なお，高速道路は駅周辺の道路網の一部であるとは考えにくく，同列に扱うのは不適切であると考えたため，道路データ内で高速道路に分類されている道路は分析対象から除いた．

駅の座標と抽出した全ての道路リンクの端点の座標と実延長のデータを用いて以下の処理を行う．

- 1) 道路リンクの端点に対して駅から近い点を道路リンクの起点，遠いほうを終点とする．
- 2) 起点と終点から道路リンクの偏角を求める．
- 3) 起点と終点の中点を道路リンクの代表点とし，駅からの距離，方角を求める．

この作業を図に落とした概略図を図-1 に示す．

これで，道路リンクの属性として，"偏角(θ)"，"実延長"，"駅からの距離"，"駅からの方角(ϕ)" が得られた． θ と ϕ の定義域は $-\pi$ から π であり，反時計回りを正とする．これら4つの道路リンクの属性のうち，本論文で紹介する手法で用いるものは偏角のみである．

(2) 方向統計学

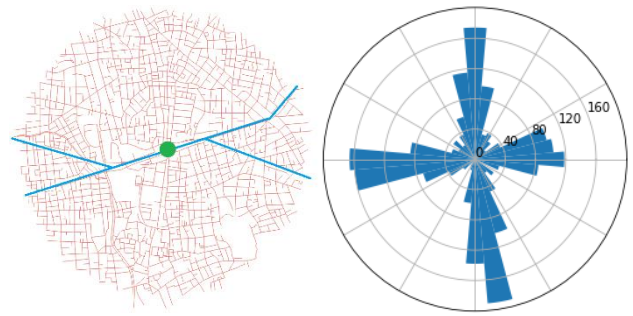
方向統計学(Directional Statistics)とは方向データ，つまり角度データを扱う統計学である．角度データは 2π の周期を持つため，通常の統計処理では様々な場面で不具合が生じるため，特殊な処理を行う．方向統計学において，ヒストグラムはバラ図(Rose Diagram)と呼ばれる特殊なグラフで代用する．通常のヒストグラムに対して左端の $-\pi$ と右端の π のデータをつなげて輪にするようなイメージである．右(東)向きの角度データが右に位置するため，直感に合うグラフになる．

(3) データの記述

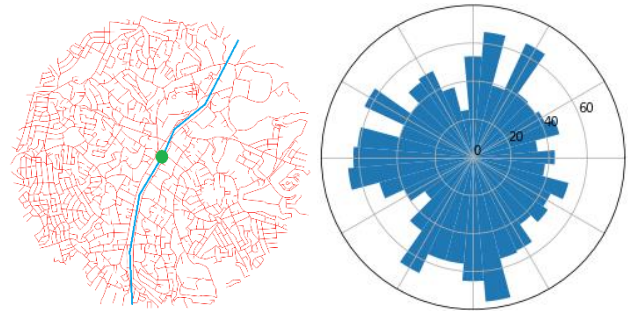
a) バラ図

駅周辺の道路リンクの偏角からバラ図を作成した．東急目黒線/大井町線・大岡山駅と小田急小田原線・玉川学園前駅周辺の道路網とバラ図を図-2 に示す．

青線は鉄道線路の線形，緑丸は駅の位置である．大岡



(a) 大岡山駅



(b) 玉川学園前駅

図-2 道路網とバラ図

山駅周辺の道路網は主に東西または南北方向の道路リンクが多いため，東西南北の四方向の偏角が多いバラ図になる．多くの駅は大岡山駅のようなバラ図になる．一方，玉川学園前駅周辺の道路リンクは方向がランダムに近く，偏角の分布は大岡山駅に比べると一様であることがわかる．バラ図を用いることで，特徴的な道路網を目で判別することができる．

b) 分布の推定

道路リンクの偏角のバラ図を目視することで道路網の特徴を理解することは可能であるが，より定量的に分析するために，偏角の確率分布を推定する．本研究では，様々な分布形状に対応できる柔軟性の高い Kato and Jones 分布³⁾ を用いて推定を行う．確率密度関数を以下の(1)に示す．

$$g(\theta) = \frac{1}{2\pi} \left[1 + 2\gamma^2 \right] \quad (1)$$

$$\frac{\gamma \cos(\theta - \mu) - \bar{\alpha}_2}{\gamma^2 + \bar{\alpha}_2^2 + \bar{\beta}_2^2 - 2\gamma\{\bar{\alpha}_2 \cos(\theta - \mu) + \bar{\beta}_2 \sin(\theta - \mu)\}}$$

ただし，

$$-\pi \leq \mu < \pi, \quad 0 \leq \gamma < 1, \quad (\bar{\alpha}_2, \bar{\beta}_2) \neq (\gamma, 0), \quad (\bar{\alpha}_2 - \gamma^2)^2 + \bar{\beta}_2^2 \leq \gamma^2(1 - \gamma^2) \quad (2)$$

を満たす。μは位置、γは集中度、 $\bar{\alpha}_2$ は尖度、 $\bar{\beta}_2$ は歪度を表すパラメータである。パラメータが四つと比較的多いため、様々な分布の形状を作る。

多くの駅のバラ図を見ると分布のピークが四つあることが確認できたため、このKato and Jones分布を混合数4で混合した分布を用いて推定を行う。便宜上この混合分布をMoKJ(Mixture of Kato and Jones)と呼ぶ。MoKJの確率密度関数を以下の(3)に示す。

$$p(\theta|\pi, \mu, \gamma, \bar{\alpha}_2, \bar{\beta}_2) = \sum_{h=1}^4 \pi_h g(\theta|\mu_h, \gamma_h, \bar{\alpha}_{2h}, \bar{\beta}_{2h}) \quad (3)$$

ここで、 π_h は全体の分布に対する h 番目の分布の割合であり、 $\sum_{h=1}^4 \pi_h = 1$ となる。推定するパラメータは各 $\pi_h, \mu_h, \gamma_h, \bar{\alpha}_{2h}, \bar{\beta}_{2h}$ であるが、そのうち π_4 は $1 - \pi_1 - \pi_2 - \pi_3$ と等しいため、推定するパラメータ数は19個である。

パラメータの推定にはモーメントを最大化するアルゴリズムを用いる。そのアルゴリズムをEMMアルゴリズム(Expectation-Moment-Maximisation algorithm)と呼ぶ。EMMアルゴリズムは尤度最大化のアルゴリズムであるEMアルゴリズムを元にしており、尤度の最大化をモーメントの最大化で代用することにより計算時間の短縮を図っている。Kato and Jones分布では、モーメントを最大化するパラメータの値と尤度を最大化するパラメータの値は概ね一致することはわかっている。しかし、それを混合したMoKJを用いた場合には、モーメントを最大化するEMMアルゴリズムの収束は理論的に保証できていないわけではない。しかし、著者の知る限りではMoKJのパラメータ推定方法は確立されていないため、ひとつの実行可能な方法としてEMMアルゴリズムを提案する。

EMMアルゴリズムの手順を説明する。最初に、パラメータの初期値を設定する。そして、サンプルデータ θ_n ($n=1,2,3,\dots,N$)が得られたとき、各データ θ_n が $h=1,2,3,4$ のどの分布から発生したデータであるのかという程度を表す負担率を以下の式(4)で計算する。(E-step)

$$\rho(z_{nh}) = \pi_h g(\theta_n) / \sum_{j=1}^4 \pi_j g(\theta_n) \quad (4)$$

ここで、 $\sum_{h=1}^4 \rho(z_{nh}) = 1$ である。負担率から新しい π_h を以下の式(5)で求める。

$$\pi_h = \sum_n \rho(z_{nh}) / N \quad (h=1,2,3) \quad (5)$$

さらに、負担率を用いてモーメントを最大化する各パラメータを以下の式(6)で計算する。(MM-step)

$$\begin{aligned} \mu_h &= \arg(\sum_n \rho(z_{nh}) e^{i\theta_n}) \\ \gamma_h &= |\sum_n \rho(z_{nh}) e^{i\theta_n} / \sum_n \rho(z_{nh})| \\ \bar{\alpha}_{2h} &= [\sum_n \rho(z_{nh}) \cos\{2(\theta_n - \mu_h)\}] / \sum_n \rho(z_{nh}) \\ \bar{\beta}_{2h} &= [\sum_n \rho(z_{nh}) \sin\{2(\theta_n - \mu_h)\}] / \sum_n \rho(z_{nh}) \end{aligned} \quad (6)$$

EMMアルゴリズムは、得られたパラメータを用いて新しく負担率を計算するE-stepと、その負担率から新しいパラメータを計算するというMM-stepを繰り返すものである。この2つの収束判定基準は、各パラメータの変化率の最大値が一定の値 r 以下になった時とし、その時点でのパラメータを推定したパラメータとする。

なお、モーメントは必ずしも凸ではなく、推定されたパラメータは初期値に依存する。そこで一定数 m のパラメータ組が得られた時点で計算を終了する。そして得られたパラメータの組の中で最も尤度の高いものを最適なパラメータとすることにした。この際、有効なパラメータ組を効率的に得るために各パラメータの値の範囲に制限を設けた。その範囲を推定中に脱した場合、その初期値における計算を終了する。ある初期値組数 N で計算をしても一定数 m のパラメータ組が得られなかった場合は推定失敗とする。たとえば、海沿いにある駅などは道路網の偏角に強い制約があり、偏角の分布のピークは明らかに4つではないため推定失敗となる。

c) 分布のクラスタリング

バラ図の分布から推定したMoKJのパラメータを用いて、駅周辺道路網の類型化(クラスタリング)を行う。その際、以下の方法によりパラメータの基準化を行った。

まず、 μ_h の値が小さいものの順に並べ、各パラメータの添え字 h を振りなおした。さらに、位置のパラメータ μ_h について、一つ目の山を基準とした山同士の間隔の値に変換し、単位をラジアンに変換した。山の大きさのパラメータ π_h についても、一つ目の山の大きさを基準とした相対的な山の大きさの値に変換した。以上の基準化によって20個のパラメータのうち μ_1 と π_1 の二つのパラメータが固定されるので、クラスタリングで用いるパラメータは18個となる。

以上に述べた一連の方法は、今回の道路データと同じように一つの道路リンクが交差点間の座標で記述されており、その両端点の座標と中心点の座標が与えられれば、データの種類や範囲によらず適用可能である。

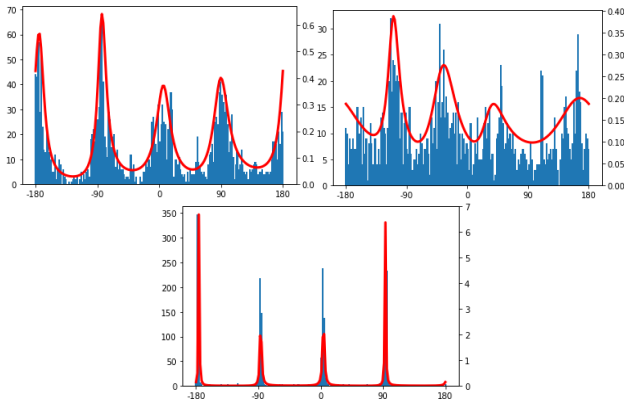


図-3 MoKJの形状(上段左:大岡山駅 上段右:津田山駅
下段:錦糸町駅)

表-1 MoKJのパラメータ

(a) 大岡山駅					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-177.0	0.9088	0.7823	0.0326	0.2476
2	-77.62	0.8442	0.7366	-0.0779	0.2701
3	7.140	0.8805	0.7096	-0.0150	0.2242
4	97.99	0.8555	0.6728	-0.1085	0.2581
(b) 津田山駅					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-100.3	0.7273	0.5994	-0.1033	0.2132
2	-17.61	0.7170	0.4605	-0.1720	0.2880
3	74.07	0.5004	0.2499	-0.2499	0.2150
4	177.8	0.7083	0.3129	-0.0840	0.2838
(c) 錦糸町駅					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-84.06	0.9465	0.9251	-0.0413	0.2476
2	4.50	0.9535	0.9292	-0.0071	0.2701
3	94.76	0.9671	0.9552	-0.0248	0.2242
4	-175.4	0.9387	0.9290	-0.0057	0.2581



図-4 津田山駅(左)と錦糸町駅(右)の道路網

3. 結果と考察

(1) 分布の推定

収束判定基準 r を0.5%, 初期値組数 M は10000組, パラメータ組数 m を20組として分布の推定を行った. 分析対象の1561駅のうち, 推定が成功したものは1434駅であった. 一例として, 大岡山駅, JR南武線・津田山駅, JR総武線/東京メトロ半蔵門線・錦糸町駅の分布を図-3に, パラメータを表-1に示す. さらに, 津田山駅と錦糸町駅周辺の道路網を図-4に示す.

このグラフにおいて, 分布の右端と左端は本来接続している. γ_h が大きいと μ_h 付近の山が高くなり, $\bar{\alpha}_{2h}$ が大きいと鋭利な分布になる. $\bar{\beta}_{2h}$ は正に大きいと右に歪み, 負に大きいと左に歪む. 特定の偏角を多く持つ錦糸町駅周辺の道路網の分布における γ_h と $\bar{\alpha}_{2h}$ は大きな値を持つ. それに対して, 様々な偏角の道路が存在する津田山駅周辺の道路網の分布における γ_h と $\bar{\alpha}_{2h}$ は小さな値になる. MoKJは尖度や歪度を表現できるため, 当てはまりのいい分布になった. 大岡山駅の $h=2$ の山の尖りや $h=4$ の山の偏りが表現されている. この分布により, 各駅周辺の道路網を定量的に比較することができる.

(2) 分布のクラスタリング

MoKJのパラメータを用いてクラスタリングを行う. 推定が成功した1434駅に対して k -means法⁴⁾を用いてクラスタリングを行う. k -means法ではクラスター数 k を所与としてクラスタリングを行うので, クラスター数 k によって結果も変わる. 以下では, $k=2,3$ の場合の結果と考察を行う.

$k=2$ の場合の各クラスターの重心のパラメータを当てはめたグラフを図-5に, パラメータを表-2に示し, $k=3$ の場合のグラフを図-6に, パラメータを表-3に示す. μ_1 は全て -135° で統一している.

$k=2$ の場合の二つのクラスターの重心を見ると, μ_h や $\bar{\beta}_{2h}$ にはほとんど違いが見られない. γ_h や $\bar{\alpha}_{2h}$ について, クラスター2のものが少し大きい値になっている. すなわち, 特定の偏角(典型的には東西南北)を向いた道路リンクが多い道路網がクラスター2に属している. 大岡山駅や錦糸町駅はこちらに属する. 対して特定の偏角を持つ道路リンクが比較的少ない駅がクラスター1に属する. 津田山駅はこちらに属する.

$k=3$ のクラスター1は $k=2$ のクラスター1に, $k=3$ のクラスター3は $k=2$ のクラスター2に似ている. $k=3$ のクラスター2は $k=2$ のクラスター1や $k=3$ のクラスター1よりも γ_h と $\bar{\alpha}_{2h}$ が小さいため, 特定の偏角を持つ道路リンクがかなり少ない駅がクラスター2に属する. 玉川学園前駅はこちらに属する.

しかし, 基準化の影響で π_1 と π_h の比である π_h' は0.5か

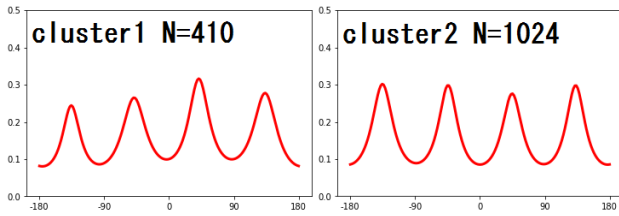


図-5 $k=2$ におけるクラスターの重心の分布

表-2 $k=2$ におけるクラスターの重心のパラメータ

(d) クラスター1					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-135.0	0.7089	0.5404	-0.0086	0.1898
2	-48.02	0.7173	0.4972	-0.0058	0.2604
3	42.15	0.7176	0.5228	-0.082	0.2822
4	132.4	0.7285	0.5083	0.0109	0.2676

(e) クラスター2					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-135.0	0.7408	0.5378	-0.0044	0.2676
2	-44.96	0.7321	0.5467	0.0081	0.2480
3	44.31	0.7317	0.5376	0.0038	0.2372
4	132.2	0.7343	0.5490	0.0071	0.2472

ら2程度の値を取り、 γ_h は0.5 から1 までの値のみしか取れないため、後者に比べて前者のほうが駅間のパラメータの差が大きくなる傾向にある。今回の k -means法ではパラメータの種類による距離の重みをつけていないため、距離の差が大きく出得るパラメータの影響がかなり大きくなってしまふ。同様の原因で1.5 から4.5程度の値を取る μ_h の影響はかなり大きい。この分析において、山の集中度、山の大きさなどの各パラメータのうち、どれが道路網の分類に強く影響する要素なのかという考察はまだ十分に行えていない。また、山の組み合わせは同じだが順番が違うような分布は別の分布とみなしてしまう。このような分布と、位置関係は同じだが各山のパラメータの組み合わせは違う分布の類似度の差異をどのように付けるかという考察も行えていない。これらの要素も反映できるような解析手法の提案は課題の一つと考えられる。

4. 結論

本研究では、各道路の偏角に着目しバラ図を作成するとともに、偏角の確率分布を推定する手法を考案した。このことにより、偏角の分布の集中度や尖度、歪度などがわかるようになり、推定されたパラメータから道路網

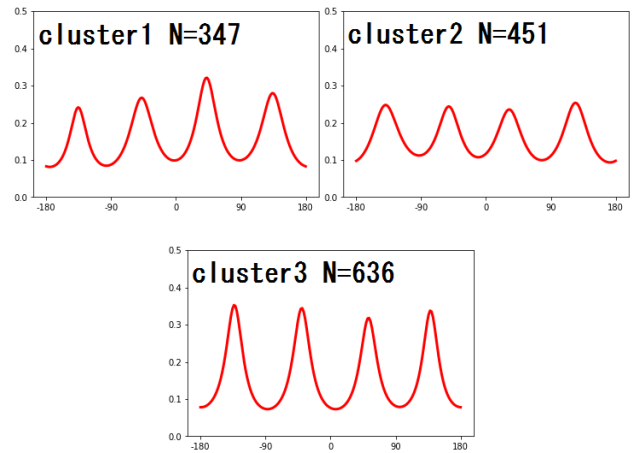


図-6 $k=3$ におけるクラスターの重心の分布

表-3 $k=3$ におけるクラスターの重心のパラメータ

(a) クラスター1					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-135.0	0.7095	0.5420	-0.0067	0.1857
2	-47.12	0.7195	0.4993	-0.0065	0.2621
3	43.26	0.7186	0.5265	-0.0084	0.2838
4	133.7	0.7295	0.5102	0.0073	0.2684

(b) クラスター2					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-135.0	0.6975	0.4559	-0.0434	0.2662
2	-52.58	0.6775	0.4744	0.0107	0.2288
3	31.37	0.6916	0.4639	0.0098	0.2417
4	123.6	0.6998	0.4735	0.0069	0.2633

(c) クラスター3					
h	μ_h	γ_h	$\bar{\alpha}_{2h}$	$\bar{\beta}_{2h}$	π_h
1	-135.0	0.7681	0.5951	0.0218	0.2624
2	-40.36	0.7681	0.5920	0.0052	0.2616
3	52.67	0.7581	0.5864	-0.0015	0.2383
4	137.6	0.7576	0.5975	0.0096	0.2378

のクラスタリングができることを示した。この手法を東京圏一都三県の1500余りの駅周辺の道路網の分析に適用し、道路網を類型化した。

本研究ではMoKJのパラメータ推定のために、EMMアルゴリズムを提案した。この方法では繰り返し計算の収束は必ずしも保証されていないため、理論的により厳密な方法を構築する必要がある。また、混合数を所与として推定を行っているが、混合数も同時に推定できる方法を構築することは今後の課題の一つである。さらに、クラスタリング手法の改善によってさらに一般性のある類型化が可能になると考えられる。

道路リンクが持つ属性のうち、道路の延長、駅からの距離と方角については、バラ図に次元を追加したシリン

ダーやトーラスといった方向統計学の概念を用いて分析できるため、別途に分析を進めている。その他、道路幅員や管理種別などを考慮することや、道路リンクのカーブを表現することで、さらに細かい分析が可能になると考えられる。駅周辺の道路網の形状の指標と、駅周辺の活動指標を比較し、道路網が駅周辺の活動にどのような影響を与えるのかという分析を行うことも今後の課題の一つであると考えている。

謝辞：本研究は科学研究費(若手研究:18K13846)の助成により行われた。

参考文献

- 1) Masucci, A. P., Stanilov, K. and Batty, M.: Limited Urban Growth: London's Street Network Dynamics since the 18th Century, *PLOS ONE*, Vol. 8, No. 8, pp.e69469, 2013.
- 2) Boeing, G.: Urban Spatial Order: Street Network Orientation, Configuration, and Entropy, Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3224723>, February 1, 2019.
- 3) Kato, S. and Jones, M.C.: A tractable and interpretable four-parameter family of unimodal distributions on the circle, *Biometrika*, Vol. 102, No. 1, pp. 181-190, 2015.
- 4) Duda, R. O., Stork, D. G. and Hart, P. E.: *Pattern Classification Second Edition*, John Wiley & Sons, Inc., pp.528-529, 2001.

(2019.3.5 受付)

ANALYSING ROAD NETWORKS BY USING DIRECTIONAL STATISTICS

Kota NAGASAKI, Wataru NAKANISHI and Yasuo ASAKURA