

大規模車両軌跡データのみを用いた 主要交差点抽出

鐘 恒毅¹・中西 航²・安田 昌平³・井料 隆雅⁴

¹学生会員 東京工業大学 環境・社会理工学院 (〒 152-8552 東京都目黒区大岡山 2-12-1)

E-mail: zhong.h.aa@m.titech.ac.jp (Corresponding Author)

²正会員 金沢大学准教授 理工研究域 (〒 920-1192 石川県金沢市角間町)

E-mail: nakanishi@se.kanazawa-u.ac.jp

³正会員 東京大学助教 大学院工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1)

E-mail: s.yasuda@civil.t.u-tokyo.ac.jp

⁴正会員 東北大学教授 大学院情報科学研究科 (〒 980-8579 仙台市青葉区荒巻字青葉 6-6-06)

E-mail: iryo@tohoku.ac.jp

車両軌跡データを分析し道路ネットワークを構築する既往研究の多くは、ナビゲーション等で用いる詳細なネットワークを対象としている。一方で、交通量配分等の用途で望まれる、主要な交差点と道路のみで構成されるネットワークはあまり扱われていない。そこで本研究では、車両軌跡データのみを用いて主要な交差点を抽出する方法を開発した。具体的には、交差点付近の軌跡データの特徴を調べたうえで、進行方向の角度分布および大きな回転角の空間分布それぞれについて、空間スキャン統計を適用した。その結果、角度分布の分散が大きい領域、あるいは大きな回転角が空間的に集積する領域が交差点として抽出された。さらに、抽出された交差点を用いたネットワークの構築を試みた。事前に地図情報を必要としない、軌跡データのみからのネットワーク生成の可能性が示唆された。

Key Words : intersection detection, GNSS trajectories, spatial scan statistics, road map generation

1. はじめに

交通ネットワーク分析においては、ベースとなる道路ネットワークデータが必要不可欠である。近年の道路ネットワークデータは精細化が進み、データ量が膨大化する傾向にある。例えば、一本の道路を車線ごとに異なる複数のリンクで表したり、簡単な交差点を4つ以上のノードと複数の短いリンクで構成したりするなど、道路の細部構造まで描く複雑なネットワークが多く見られる。日本で標準的に利用されている全国デジタル道路地図データベース (DRM) では、総延長約 100 万 km の道路リンクが約 596 万本のリンクに分けられている。その結果、一本のリンクの平均長は 200m 未満となっている。このような詳細な道路ネットワークはカーナビゲーションなどの用途には有用であり、活用されている。

しかしながら、このような詳細データを交通量配分やネットワーク脆弱性分析などの用途に直接使うと、様々な問題が生じる。特にネットワークデータと車両軌跡データ両方のデータ量が膨大な場合、計算量が上がると、いわゆる「次元の呪い」という計算時間と計算誤差がともに増大する現象が生じやすい。これらの問題を解決するために、道路ネットワークを簡略化する手法

がいくつか開発された。例えばネットワーク上の利用者均衡配分では、ネットワークアグリゲーションと呼ばれる道路ネットワークを簡略化する方法が用いられている。これは、詳細なネットワークデータの解像度を必要に応じて粗くする操作といえる。

一方、2000 年に GPS の選択利用性 (Selective Availability) が解除されて以降、民間が使える GPS/GNSS データの精度が大きく改善され、車両軌跡データが容易に使えるようになった。たとえば、現在の日本では ETC2.0 のように車両軌跡データを迅速かつ高精度に収集できる環境が整備されている。このようにして作られた車両軌跡データは交通ネットワーク分析の分析対象として活用されるだけでなく、近年では道路ネットワーク作成にも利用されている。たとえば、一本の道路における車線の分離¹⁾ や複雑な形状の交差点の構造表現など、細部の道路ネットワーク構造に着目して、ナビゲーションに使えるネットワーク生成を目指している²⁾。ここでは、GNSS データから直接道路ネットワークを作成または更新する可能性と計算の容易性が示されている。

以上を踏まえると、大規模な車両軌跡データから、経路配分などの用途に適した解像度の道路ネットワークを直接生成することは、これまでほとんど行われてい

ない。本研究ではこの目標の第一歩として、車両軌跡データのみから経路選択が多数行われる場所にあたる主要交差点を抽出することを目的とする。なお、車両軌跡データと道路ネットワークの両方の情報を用いてこの目的を達成する方法は、著者らにより別に提案されている³⁾。これに対し、本研究は真に車両軌跡データのみを手がかりに交差点抽出を試みる点で異なる。

さて、抽出された交差点は道路ネットワークにおいてノードで表現される。そのノードの領域が定めれば、大規模車両軌跡データを用いて事後的にノード間を結ぶ適当なリンクを作成することで、半自動的に道路ネットワークが構築できる。これらは本研究のあとに続く手順であり、次の機会に報告したい。なお、本研究で扱う交差点は一般道の通常の交差点のように平面上で車線が交差している場所を想定している。立体交差やジャンクションなど複雑な交差点についての処理も、次のネットワーク構築の段階で別途対応する予定である。

以降では、2章で本研究で交差点抽出に用いる空間スキャン統計を説明する。これは、交差点には車両軌跡データがある種の集積として存在していると考えられるためである。具体的には、3章で説明するデータの方角および回転角を統計量に設定する。これは、空間スキャン統計における新しい応用である。それらを受けて、4章で具体のモデル化を、5章で実データへの適用を行う。

2. 空間スキャン統計

(1) 手法の概略

空間スキャン統計は、空間上における点事象のクラスターを検出する手法の一つである。この手法はKulldorf(1997)より開発され、事象の発生率が全体領域における事象発生率と有意に差がある小領域をクラスターとし、尤度比を統計指標として小領域がクラスターであるかどうかを評価する。

一般的な空間スキャン統計では、まず全体の点事象が分布している領域を探索領域 G とし、探索領域に存在している点事象 i はある発生率 $P_{\text{all}}(i)$ にしたがって分布していると仮定する。点事象の空間分布を扱うとき、点事象の分布はポアソン過程に従うと仮定することが多い。次に、クラスターを検出するためクラスターの候補領域であるスキャン領域 Z を設定する。スキャン領域は探索領域の内部に存在しなければならない。探索領域と同様に、スキャン領域内部にある点事象 i もある発生率 $P_{\text{in}}(i)$ にしたがって分布しているとする。

ここで、あるスキャン領域 Z に対して、クラスターであるかどうかを尤度比指標で評価する。対立仮説 H_1 は「スキャン領域内の点事象の発生率がスキャン領域

外の点事象の発生率と異なる」と、帰無仮説 H_0 は「スキャン領域内の点事象の発生率がスキャン領域外の点事象の発生率と一致する」とする。スキャン領域内の事象数を n_{in} 、スキャン領域外的事象数を n_{out} 、探索領域内の全事象数を $n_{\text{all}} = n_{\text{in}} + n_{\text{out}}$ とし、また探索領域外の点事象 j の発生率を $P_{\text{out}}(j)$ とするとき、スキャン領域 Z における対立尤度 L と帰無尤度 L_0 はそれぞれ次式のように計算する。

$$L = \prod_{i=0}^{n_{\text{in}}} P_{\text{in}}(i) \times \prod_{j=0}^{n_{\text{out}}} P_{\text{out}}(j) \quad (1)$$

$$L_0 = \prod_{i=0}^{n_{\text{all}}} P_{\text{all}}(i) \quad (2)$$

尤度比 λ は $\lambda = L/L_0$ と定義され、尤度比が大きければ大きいほど、スキャン領域 Z がクラスターである確率が高い。なお、実計算中では尤度比に対数をとった対数尤度比 $\log \lambda$ を使うことが多い。

ただし、実応用では差があるだけではなく、ほかの条件も参考にしてクラスターを判断することが多い。例えば点が空間的に集中している場所をクラスターとすると、クラスター領域内の点の発生率は領域外と異なっており、かつ領域外より大きい発生率を持っている。ただし、何も条件を設けない場合、領域外より有意に小さい発生率を持っている領域もクラスターと判断される。この問題に対処するため、指示関数 I を導入して、ある条件が満たされている場合のみ $I = 1$ とし、満たされていない場合では $I = 0$ とする。一例として、点事象の密度が高いクラスターを探したい場合、指示関数 I は次のように定義できる。なお、スキャン領域内外の面積をそれぞれ $c_{\text{in}}, c_{\text{out}}$ とする。

$$I = \begin{cases} 1 & \left(\frac{c_{\text{in}}}{n_{\text{in}}} > \frac{c_{\text{out}}}{n_{\text{out}}} \right) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$

したがって、スキャン領域 Z の尤度比関数 $\lambda(Z)$ は次式となる。

$$\lambda(Z) = \frac{L(Z)}{L_0(Z)} \times I \quad (4)$$

探索領域内に設定されている全てのスキャン領域 Z に対して、尤度比 $\lambda(Z)$ を計算し、尤度比が高い順にスキャン領域を並べ、尤度比が最大であるスキャン領域 Z を最大尤度比クラスター (Maximum likelihood cluster, MLC) とする。なお、クラスターの有意性は Monte Carlo 法でランダムな分布にしたがって点事象を生成する場合の最大尤度比と比較することで判断する。あるクラスター候補の尤度比が Monte Carlo 法で計算された最大尤度比より小さければ、当該クラスター候補を

棄却する。また、領域に重複があるクラスターについては、尤度比が大きいものから順に残し、そのクラスターと重複する領域を含むクラスターは採用しない。

(2) 手法の拡張例

Kulldorff は空間スキャン統計の手法を提案し、円や楕円などの単純な形状のクラスター検出における一般的な手法を開発したが⁴⁾、スキャン領域の形状設定自体は単純であり、複雑形状のクラスター検出には適していない。そこで、糟谷ら (2012) はスキャン領域を複数の小領域の組み合わせと考えることで柔軟形状のクラスター検出手法を提案した。この手法ではまず点事象が存在している小領域を抽出し、それらの中の隣接した複数の小領域をいくつかグループに分け、グループごとに小領域の全て可能な組み合わせをスキャン領域とし、尤度比を計算する。なお、小領域は一般的な行政区画と設定している。次に、尤度比が大きいスキャン領域をクラスターとして、隣接したものを結合し再び尤度比で評価する⁵⁾。また、井上 (2013) はこれをユークリッド空間上に適用できるようにし、スキャン領域の候補である小領域をリンクとすることでネットワーク上でのクラスター検出に拡張した⁶⁾⁷⁾。さらに、井上ら (2016) は形状複雑度を導入し、一般的な尤度比関数の代わりに特定形状のクラスター検出ができるようになった⁸⁾。

(3) 既往研究での応用例

Kulldorff (1995) は、ポアソン分布を仮定した尤度比検定式を導出した⁴⁾。また、Kulldorff (1997) は、探索領域内に分布している点事象がベルヌーイ分布に従うときの尤度比検定式を導出した⁹⁾。これらの式を用いて、Kulldorff (1997) は事象の空間分布に着目して、ノースカロライナの乳幼児突然死症候群 (SIDS) ケースの集積を調べた⁹⁾。同様にネットワーク上の空間分布に着目して、さらに時間方向に拡張したのは Shiode (2011) で、街路上犯罪案件の集積検出をした¹⁰⁾。これらの研究はいずれも地点の分布に関する分析である。

さらに Kulldorff et al. (2009) は、正規分布を仮定した尤度比検定を提案した¹¹⁾。これを用いて、渡邊ら (2011) は空間分布だけでなく、地価の分布を正規分布とし不動産価格水準や取引位置の時空間集積を抽出した¹²⁾。しかし、経緯度情報をはじめとした経路軌跡データの空間分布や角度分布に関して、空間スキャン統計を適用する研究はほとんど存在しない。

(4) 本研究での考え方

本研究では、車両軌跡データが交差点で何らかの集積をしていることを仮定する。対象の道路領域全体を

探索領域として、空間スキャン統計を用いて集積の範囲を定め、交差点領域を集積として抽出する。詳細は次章で述べるが、ごく単純に考えると、

- 交差点は交差点以外よりも車両軌跡データ自体が多い
 - 交差点は交差点以外よりも回転する車両が多い
- といったことが想定される。これらを集積と捉えるために、車両軌跡データに適した統計量の設計を行っている。また、現実の交差点の形状はそれぞれ異なるため、スキャン領域の形状設定にも工夫が必要である。以上より本研究では、複数の統計量の分布における尤度比関数を設定し、それぞれについて柔軟な形状のクラスター検出を試みる。

3. 統計量の作成

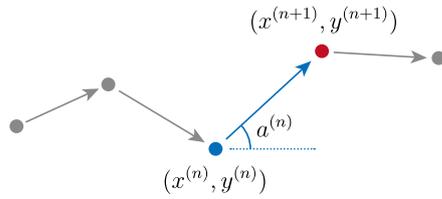
(1) 車両軌跡データ

本研究では大規模車両軌跡データの一例である ETC2.0 のデータを用いて主要交差点の抽出を行う。なお、多くの車両軌跡データは ETC2.0 と同様に緯度経度の時系列データとして与えられるため、以下の方法は一定程度の汎用性を有する。

ETC2.0 は高度道路交通システム (ITS) の一種であり、車両走行中の位置情報を記録することで車両の経路情報の把握ができるシステムである。2011 年から順次整備され、2020 年時点で約 493 万台の車両が ETC2.0 の車載器を搭載し、また高速道路に約 1700 カ所、直轄国道には約 2000 カ所の路側機が整備されている。ETC2.0 データには車両の走行履歴データと挙動履歴データが含まれている。走行履歴データは時刻、車両の位置 (緯度および経度)、速度等を指し、車両の走行距離が 100m もしくは 200m 増加した時、または車両の進行方向が 22.5 度もしくは 45 度変化した時に記録される。走行履歴データは車両の前後左右の加速度とヨー加速度などを指し、車両の加速度が 0.25G 以上またはヨー角速度が $\pm 8.5 \text{ deg/s}$ 以上変化した時に記録される。これらのデータは平常時車両に搭載されている ETC2.0 車載器に保存している。車両が路側機の下を通過すると、車載器にプローブ情報データは吸い上げられ、プローブサーバーにより統合・集計される。

(2) 角度データ統計量

本研究では、角度情報である方位角と回転角について、リンク (通常の道路部分) とノード (交差点部分) での分布の違いを利用し、ノードすなわち交差点を検出する。ただし、ETC2.0 データのような一般的な車両軌跡データにはそれらの情報が含まれていないため、角度情報をあらかじめ定式化し、経緯度情報から算出す

図-1 方位角 $a^{(n)}$

る必要がある。

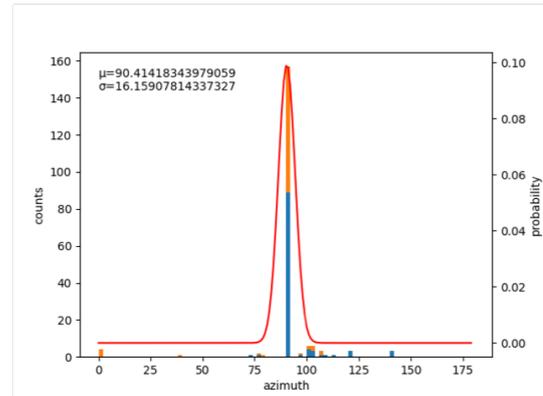
a) 方位角

方位角は車両の進行方向を表す角度指標である。図-1に示すように、ある車両軌跡上のデータ d_n の方位角 $a^{(n)}$ は、 d_n との次の時点で記録された d_{n+1} の位置により定義される。平面直角座標系上にある d_n の座標を $(x^{(n)}, y^{(n)})$ とすると、 d_n の方位角 $a^{(n)}$ は次の式より計算される。

$$a^{(n)} = \tan^{-1} \frac{y^{(n+1)} - y^{(n)}}{x^{(n+1)} - x^{(n)}} \quad (5)$$

右左折や車線変更時などの回転行動を伴う場合以外では、道路上を走行している車両の方位角は走行道路の角度と一致していると想定される。そのため、方位角の角度分布は、交差点付近とそれ以外とは以下のように異なる特徴を示すと考えられる。まず、道路リンク上または交通量が少ない細街路との小型交差点付近では、多くの車両軌跡は主要道路であるリンクに沿う。したがって、車両軌跡データのほとんどは道路リンクの向きと同じ方位角を持っている。このとき、方位角の角度分布は単峰で、角度分布の平均が道路リンクにおける車両の進行方向と一致していることが期待される。なお、主要道路が一方通行でない限り、道路リンクは進行方向の差が 180 度の二つの方向を持つ車線から構成される。このとき、両車線上の軌跡データの方位角には互いに概ね 180 度の差が存在する。ある狭い範囲に着目して、これらの方位角の角度分布をヒストグラムに記述すると、角度の差が 180 度の 2 つのピークが生じる。これは平均の差が 180 度の 2 つの単峰型分布の混合分布とも考えられる。一方、交差点ノードは 2 つ以上のリンクが交わる場所である。そのため、交差点付近の方位角の角度分布は、この交差点を通る全てのリンクの角度分布の混合分布になると考えられる。たとえば、一般的な十字路では、2 組の差が互いに 180 度の計 4 つのピークが観測される。

ここで、方位角の角度分布を単峰の分布（たとえば正規分布）として扱ううえで、以下 2 つの前処理が必要である。まず、180° 以上の方位角について、180° を引いて 0° ~ 180° の範囲におさめるようにする。これは道路を無向線分として考え、上述の進行方向による影響をなくすためである。この処理を行うと、リンク上の



青と橙が方位角データ。橙は 180° 以上の方位角から 180° を引いた値。赤実線は正規分布で近似した場合の密度関数。次図も同様。

図-2 リンクと思われる領域の方位角の角度分布の例

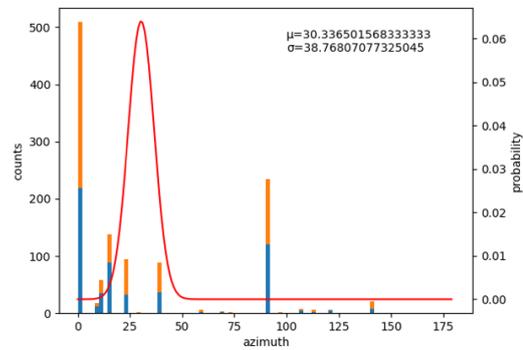


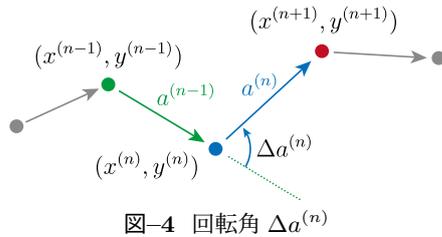
図-3 ノードと思われる領域の方位角の角度分布の例

方位角は単峰分布となり、たとえば次章で述べるように正規分布を仮定すれば尤度比関数の設定が簡単になる。また、車両の進行方向である方位角は走行している道路自身の方向に強く影響される。そして、リンク上の領域の角度分布のピークは、当然この地点の道路方向とほぼ一致する。いま知りたいことは、方位角がある方向に集中しているかどうかであり、方位角の平均値そのものではない。そこで、道路方向の影響をなくすために、着目している領域内のデータについて、角度平均が 0° になるように角度座標を定義し直す。

上記の処理を行うことで、リンクとノード付近の方位角の角度分布の違いを明らかにし、分散などの統計量で捉えることができる。このような角度分布の違いを利用することで、車両軌跡データから方位角情報を用いてリンクとノードを区別することができると考えられる。なお、以上で述べたリンクとノードと違いの一例を図-2 と図-3 に示す。

b) 回転角

回転角は車両の進行方向、すなわち方位角の変化量を表す角度指標である。図-4 に示すように、ある車両軌跡上のデータ d_n の回転角 $\Delta a^{(n)}$ は、 d_n の方位角 $a^{(n)}$

図-4 回転角 $\Delta a^{(n)}$

とその前の時点で記録された d_{n-1} の方位角 $a^{(n-1)}$ により定義される。また、一連の軌跡データのうち最初のデータ d_1 について、前の時点にあたるデータは存在しないものの、車両の起動は回転とみなされないため、 d_1 の回転角は $\Delta a^{(1)} = 0$ とする。したがって、 d_n の回転角 $\Delta a^{(n)}$ は次の式より計算される。

$$\Delta a^{(n)} = \begin{cases} 0 & (n = 1) \\ |a^{(n)} - a^{(n-1)}| & (\text{otherwise}) \end{cases} \quad (6)$$

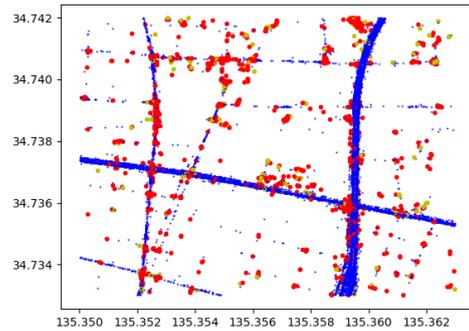
車両が直進している間は、進行方向が途中で変わらないため、軌跡データは小さい回転角を持つ。また、前述の通り、一般的な軌跡データは一定または不定な時間間隔で離散的な点として記録される。したがって、長く半径の大きなカーブ上の軌跡は方位角の変化量が小さいことが多く、回転角は小さい値となる傾向がある。一方、軌跡上に右左折行動や急カーブが発生している場所では方位角の変化率が大きく、隣り合う両データの方位角に大きな差がつくことが一般的であり、大きな回転角が計算される傾向がある。したがって、左折・右折が頻繁に発生している交差点では大きな回転角を持つデータが集中していることが予測される。特に、冒頭に述べた本研究の抽出目的である、経路選択が多数行われる交差点では、右左折をする車両が多いため、リンク上や抽出対象としない交差点よりも大きな回転角を持つデータが集中しているはずである。この想定妥当性について、一例として図-5 にデータの状況を示す。

なお、測位誤差により経路から大きく離れているデータについても計算上大きな回転角を持つ。ただし、このような測位誤差はランダムかつ低確率に発生するため、交差点の場合ほど集中していないと考える。

4. モデル化

(1) 探索領域の設定

2章で述べた空間スキャン統計では、データが探索領域内に分布していると仮定する。ただし、ノードは少なくとも道路内に存在することが明らかだが、事前に道路ネットワーク情報が与えられていない場合には、探索領域をなんらかの方法であらかじめ決定しなければならない。そこで、車両軌跡データは車両が走ること



青が回転角 22.5 度以下、黄が回転角が 22.5 度以上 45 度以下、赤が回転角が 45 度以上のデータ

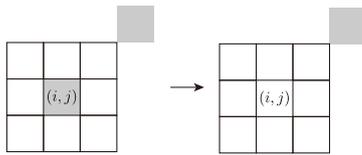
図-5 大きい回転角のデータの一例

ができる領域にしか存在しないことを利用し、探索領域を車両が走行できる領域と定義したい。

もちろん、車両はあらゆる地表面を走っていて、対象区域全体を探索領域と設定する考えもありうる。しかし、この方法を採用すると、交差点ではなく単に交通量が多い道路らしき領域全体が集積として抽出される傾向がある。これは、空間スキャン統計における集積は尤度比で判断されるため、道路と非道路の差が交差点とリンクの差よりもはるかに大きいためであると考えられる。したがって、非道路領域の影響を避けるため、本研究では道路領域と考えられる領域のみを探索領域とする。具体的には、車両軌跡データが存在している領域を道路領域と定義する。

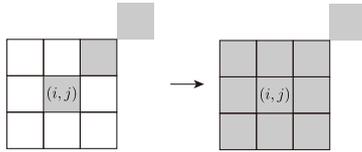
以降では、道路領域や交差点などの形状を、対象領域を分割するグリッドによって定める。グリッドは一般的な設定と同様に正方形とする。その 1 辺の長さは、交差点抽出という目的を鑑みて道路幅程度とする。本研究では経度が 0.083° 、緯度が 0.125° で定義されるおよそ 50m 四方の領域とした。各グリッドには異なる数のデータが入り、それは当該グリッドを通過する交通量に相当する。本研究では一定以上の交通量が見込まれる部分のみに着目する、すなわちデータ数がある閾値以上のグリッドを道路領域として抽出する。そのうえで、単純にグリッド内に一定数以上のデータが存在するところをすべて探索領域とすると、ノイズにより余分な領域が発生したり、逆に道路の一部が断続的になったりする問題が確認された。これは、GNSS の誤差やデータの取得間隔などデータ自身の原因によるものが多いと考えられる。これをそのまま探索領域として使うと、スキャン領域が余計に発生して効率が低くなったり、探索領域の欠損でスキャン領域の形状が正しくならなかったりすることがありうる。

そこで、より適切な探索領域の形状を構成するため、マスキングの手法を用いて、抽出された道路領域に対



灰色が道路領域， (i, j) 周りの 8 つのグリッドがマスク．マスクに入っている道路領域の数が 0 のため (i, j) を道路領域から外す．

図-6 道路領域ノイズの消去



灰色が道路領域， (i, j) 周りの 8 つのグリッドがマスク．マスクに入っているグリッドをすべて道路領域にする．

図-7 道路領域の拡張

してノイズの消去や領域の拡張を行い，道路領域の形状を補正する．マスキングとは，あるグリッドを中心グリッドと設定し，中心グリッド周囲の特定形状のグリッドの集合をマスクとし，中心グリッドまたはマスクの状態から，中心グリッドまたはマスクの状態を変える手法である．

具体的には，以下の 2 種類を行う．1 つ目は，ノイズの消去である (図-6)．中心グリッド周囲の 8 グリッドをマスクと設定し，抽出された道路領域に属する各グリッドに対してマスクの属性状態を調べる．マスクがすべて非道路状態であれば，その中心グリッドは孤立しているとし，中心グリッドの状態を非道路にする．2 つめは，道路領域の拡張である (図-7)．中心グリッド周囲の 8 グリッドをマスクと設定し，1 つめのノイズの消去後の道路領域に属する各グリッドに対して，マスクにあるグリッドをすべて道路状態にする．なお，このプロセスは，実際のデータの状態をみながら複数マスクを設定しマスキングを複数回行うことも考えられる．

(2) スキャン領域の設定

柔軟形状のクラスター検出では，検出単位である小領域の組み合わせを考え，可能な小領域の組み合わせを全て列挙し，それぞれの尤度比を計算し統計的にクラスターであるかどうかを判断することが一般的である．しかし，組み合わせの種類は小領域の個数が多くなるにつれて指数的に増大する．たとえば，中心グリッドが固定の半径が 1 の 3×3 のグリッドを考えると，中心グリッドを含む可能な小領域の種類は，周囲 8 グリッドを含むか含まないかの $2^8 = 256$ 通りもある．半径を大きくするとさらに組み合わせ数が多くなる．その結果試行回数が増え，計算時間が現実的でなくなる．

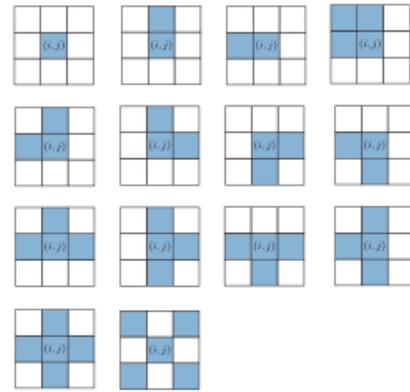


図-8 スキャン領域の種類

したがって，あらかじめスキャン領域の大きさや形状複雑度などを定め，形状制限を設定する必要がある．

ここでは，検出したい交差点の直径を 150m 以下とし， 3×3 のグリッドに収められるような形状について考え，スキャン領域として用いる形状や大きさを図-8 の 14 種類に決定する．まず，道路同士が交差する場所として一般的なノードである交差点では，横断歩道を交差点の境界として見ると，正方形または長方形となることが多い．このタイプとして 4 種類のスキャン領域を考える．これは一番簡単な交差点の形状である．交差する道路の道路幅を考え，ここでは 1×1 ， 2×2 および 1×2 の 3 種の形状のスキャン領域が考えられる．なお， 1×2 は横と縦の 2 種類がある．また，大型道路同士の交差点は太い十字となることもある．一般的な十字交差点は + 型の 5 つのグリッドからなる．これは道路の向きによって回転しうるため 2 種類考えられる．さらに，交差点の中心とグリッドとの位置関係によって，形状が T 型や L 型となることもある．これに 8 種類考えられるため，合計 14 種類が設定される．

(3) 尤度比

前章で設定した回転角については空間分布指標を，方位角については角度分布指標をそれぞれ用いて尤度比の算出を行う．以降で順に説明する．

a) 空間分布指標

回転挙動をしているときに記録された回転角が大きいデータは交差点付近に集積していると仮定される．これらのデータが集積している領域を，空間分布指標で抽出する．

いま，帰無仮説を「回転挙動は道路領域内にランダムに発生する」とし，回転角の大きいデータが道路領域内でポアソン分布にしたがって分布しているとする．スキャン領域内にあるデータの個数を n_{in} ，スキャン領域外にあるデータの個数を n_{out} ，探索領域全体にあるデータ数を $n_{all} = n_{in} + n_{out}$ とすると，このときの尤

度比関数は

$$\lambda = \frac{\left(\frac{n_{in}}{c_{in}}\right)^{n_{in}} \left(\frac{n_{out}}{c_{out}}\right)^{n_{out}}}{\left(\frac{n_{all}}{c_{all}}\right)^{n_{all}}} \quad (7)$$

である。

また、スキャン領域内に回転角の大きいデータが周囲よりも有意に多いか少ないかはデータの密度で判断できる。交差点抽出にあたっては密度の大きいスキャン領域のみを残す必要があるため、指示関数 I は次のように定義する。

$$I = \begin{cases} 1 & \left(\frac{n_{in}}{c_{in}} > \frac{n_{out}}{c_{out}}\right) \\ 0 & (\text{otherwise}) \end{cases} \quad (8)$$

なお、本研究は同じ面積の小領域の組み合わせをスキャン領域としたため、計算の便宜上、スキャン領域内外の面積 c_{in} と c_{out} を計算することなく、スキャン領域内外が含まれる小領域の個数 g_{in}, g_{out} で代用している。

b) 角度分布指標

交差点のようなノード上と交差点ではないリンク上の方位角の角度分布が異なり、ノード上では角度分布の分散が大きいと仮定する。この角度分布の違いを用いて、分散が大きい分布が集中している領域を抽出する。

いま、帰無仮説を「すべての道路領域内の方位角は同じ分布に従う」とする。ここで、角度の分布を各グリッドに対して平均が 0 の正規分布と仮定する。帰無仮説のもとで尤度が最大となる場合の平均と標準偏差は、それぞれ $\mu = 0, \sigma^2 = \sum_i (\mu - x_i)^2 / N = \sum_i x_i^2 / N$ である。したがって、帰無尤度 L_0 は次式によって計算される。

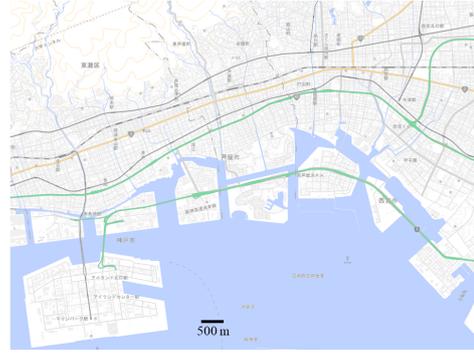
$$L_0 = \prod_{i=0}^{n_{all}} \frac{1}{\sqrt{2\pi}\sigma_{all}} \exp\left(-\frac{x_i^2}{2\sigma_{all}^2}\right) \quad (9)$$

帰無尤度 L_0 の対数尤度 $\ln L_0$ は次式によって計算される。

$$\ln L_0 = -N \ln(\sqrt{2\pi}) - N \ln(\sigma_{all}) - \sum_i \frac{x_i^2}{2\sigma_{all}^2} \quad (10)$$

対立仮説は、「集積内外のデータはそれぞれ分散が $\sigma_{in}, \sigma_{out}$ と異なる分散の正規分布に従う」とする。なお、探索領域やスキャン領域を構成するすべての小領域内は平均を 0° に合わせたため、集積内外とも正規分布の平均は $\mu = 0$ で一致する。このとき対立尤度 L は次式によって計算される。

$$L = \prod_{i=0}^{n_{in}} \frac{1}{\sqrt{2\pi}\sigma_{in}} \exp\left(-\frac{x_i^2}{2\sigma_{in}^2}\right)$$



国土地理院: 地理院地図に加筆.

図-9 対象領域の地図

$$\times \prod_{i=0}^{n_{out}} \frac{1}{\sqrt{2\pi}\sigma_{out}} \exp\left(-\frac{x_i^2}{2\sigma_{out}^2}\right) \quad (11)$$

また、角度分布については交差点ではリンクより大きな分散を持つ方位角分布となることから、指示関数 I は次のように定義する。

$$I = \begin{cases} 1 & (\sigma_{in} > \sigma_{out}) \\ 0 & (\text{otherwise}) \end{cases} \quad (12)$$

5. モデルの適用

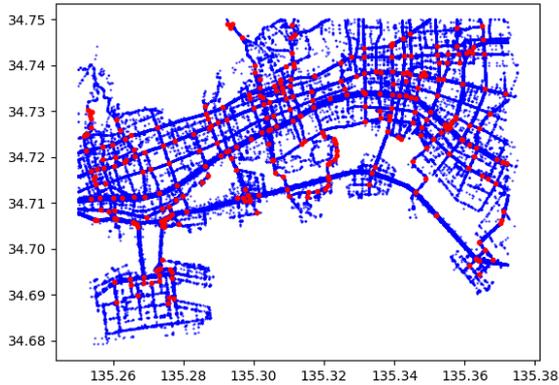
神戸付近の 1 個の 2 次メッシュにおける 2018 年 6 月 6 日の 1 日分の ETC2.0 データを用いて、提案した空間分布指標と角度分布指標の 2 種類の方法について、交差点の抽出を行った。対象領域の実際の地図を図-9 に示す。

まず、グリッド内のデータ数が 10 以上のグリッドを道路領域候補として抽出した。その後、前述のマスクング等の形状処理をした後のグリッドの集合を道路領域とする。抽出された道路領域は 9,114 個のグリッドからなり、467,248 個のデータが含まれている。

(1) 空間分布指標

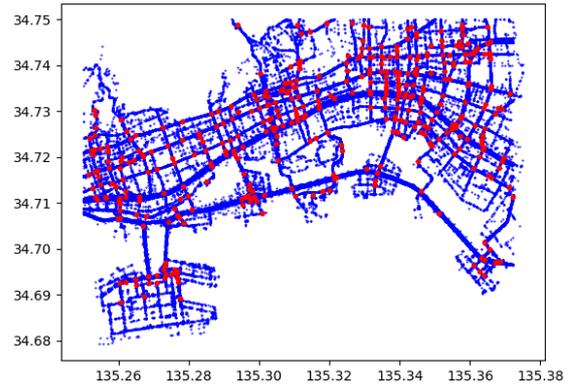
空間分布指標では、大きな回転角を持つデータが空間的に集積している領域を交差点領域として抽出する。ここで、ETC2.0 の集積条件である 22.5° 以上の回転角を持つデータを大きな回転角を持つデータとした。なお、ETC2.0 では累積角度変化でデータの集積をしており、角度変化で集積されたデータ一つひとつは必ずしも大きな回転角を持っているわけではない。よってここでは 3 章で説明した定義にしたがって回転角を計算した。方位角が 22.5° 以上の大きな回転角を持つデータは 40,526 個あり、全体の 8.67% を占めている。

Monte Carlo 法より計算された最小最大尤度比は 12.75 であった。すべてのスキャン領域のうち、尤度比



青が車両軌跡データの集合，赤が抽出された交差点。
縦軸は北緯，横軸は東経。

図-10 空間分布指標で抽出された交差点の分布



青が車両軌跡データの集合，赤が抽出された交差点。
縦軸は北緯，横軸は東経。

図-11 角度分布指標で抽出された交差点の分布

が最小最大尤度比より大きい条件が満たされたスキャン領域は 24,262 個である。また、重複がある領域を除外するプロセスを経て、最終的にクラスターとして抽出された交差点は 312 個である。1 個の交差点の大きさの平均は 3.39 グリッドであり、そのうち形状が L 型のクラスターの個数をもっとも多い。

大きな回転角を持つデータの空間分布指標で抽出された交差点を図-10 に示す。視覚的には、主要な道路同士の交差点に見えるところが抽出されている。特に、対象領域の西側にある摂津本山付近の住宅街付近のような、道路網の密度が大きくなり、格子状となっている領域に存在する交差点が正しく抽出できている。一方で、道路網の密度が大きい場所では複数の交差点が 1 つとみなされることがあった。これは、道路網の密度が大きくなると、交差点の密度も大きくなるためである。例えば、対象領域の芦屋駅付近では、道路の間隔がおおよそ 100m である。このとき、いま設定している 50m×50m のグリッドでは交差点の中心が 2 グリッド分しか離れていない。すると、両交差点が隣接する 2 つのグリッドに入り、同一の交差点に統合されてしまう。このように、道路網の密度が大きい場所ではグリッドの取り方によって交差点同士の境界が不明瞭となることがあり、空間分布指標の手法では検出しにくくなると考えられる。

また、交差点ではない道路の急カーブ部分も抽出されたことが確認できる。これは、急カーブ部分にも交差点と同様に車両回転行動が発生し、大きな回転角を持つデータが多数観測されるためである。交差点とカーブの区別は、車両軌跡を用いてノードに接続するリンクの本数を確かめれば容易に可能であり、事後的な処理を行うことで対処する予定である。

(2) 角度分布指標

角度分布指標では、方位角に大きな分散を持つデータが空間的に集積している領域を交差点領域として抽出する。Monte Carlo 法より計算された最小最大尤度比は 25.42 であった。すべてのスキャン領域のうち、尤度比が最小最大尤度比より大きい条件が満たされたスキャン領域は 15,241 個である。また、重複がある領域を除外するプロセスを経て、最終的にクラスターとして抽出された交差点は 428 個である。1 個の交差点の大きさの平均は 3.09 グリッドであり、抽出されたクラスターの形状はほとんど L 型である。

大きな方位角の分散を持つデータの角度分布指標で抽出された交差点を図-11 に示す。視覚的には、空間分布指標の場合と同様に、主要な道路同士の交差点に見えるところが抽出されている。ただし、空間分布指標よりも角度分布指標を用いたほうが多くのクラスターが抽出された。図-10 と図-11 を比べると、角度分布指標で多めに抽出されたクラスターは主にインターチェンジ付近にあることがわかる。

インターチェンジ付近で交差点領域が多く抽出された原因として、向きが異なる複数のリンクが 1 つのスキャン領域内に存在することが考えられる。一般道と異なり、インターチェンジは大規模かつ複雑な構造を持つため、必ずしも平面的には交差していない、方向の異なる道路が小さい領域内に複数存在するケースがある。したがって、回転行動がなくても、1 つのグリッドに 2 つ以上の方角を持つデータが存在しうる。その結果、リンクの角度分布とは有意に異なる領域とみなされ、交差点として抽出された。一方で、空間分布指標の場合に誤って抽出された急カーブ部分は、角度分布指標の場合には抽出されにくくなったことも確認できる。

6. 結論

本研究では、ETC2.0 のような車両軌跡データの経緯度情報のみを用いた交差点抽出手法を提案した。まず、交差点付近での車両軌跡データの特徴分析を行い、大きな回転角を有するデータの空間分布および方位角の角度分布の分散に、交差点付近特有の傾向があることを明らかにした。そのうえで、それらの特徴を統計量として定義することで、空間スキャン統計を利用した交差点抽出手法を考案した。この手法を ETC2.0 データに適用し、神戸付近の交差点抽出を行った。その結果、方位角や回転角を用いれば、既存の道路ネットワーク情報を用いなくても車両軌跡データのみから交差点抽出が可能であることが示された。

今後の課題と展望は以下である。まず、手法の精度向上である。交差点ではない場所が誤抽出されたり、抽出された交差点領域が実際の場所とややずれたりする問題の解決が望まれる。特に、急カーブのように本研究の帰無仮説の設定に合致しないものを、本来的な意味に即して帰無仮説に含める方法の検討が必要である。また、異なるデータセットへの適用を通して手法の有用性を高めることが、これらの課題解決にも役立つと考えている。この際には、現状では視覚的判断に留まっている精度評価手法自体も検討する必要がある。そのうえで、抽出された交差点に基づいてネットワークを構築し、交通量配分等の用途に適用していくことを目指したい。

謝辞: 本研究の一部は、国土交通省「道路政策の質の向上に資する技術研究開発」における研究「ETC2.0 データの活用と評価を通じた次世代 ETC の基本設計提案」(代表:井料隆雅)の協力のもとで実施した。また、科学研究費(若手研究: 18K13846 および基盤研究(B): 20H02267)の助成により行われた。

参考文献

- 1) Uduwaragoda, E. R. I. A. C. M., Perera, A. S., and Dias, S. A. D.: Generating lane level road data

- from vehicle trajectories using kernel density estimation, *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pp. 384–391, 2013.
- 2) Tang, J., Deng, M., Huang, J., and Liu, H.: A novel method for road intersection construction from vehicle trajectory data, *IEEE Access*, Vol.7, pp.95065–95074, 2019.
- 3) Yasuda, S., Iryo, T., Sakai, K., and Fukushima, K.: Data-oriented network aggregation for large-scale network analysis using probe-vehicle trajectories, *IEEE ITSC*, pp. 1677–1682, 2019.
- 4) Kulldorff, M. and Nagarwalla, N.: Spatial disease clusters - detection and influence, *Statistics in Medicine*, Vol.14, No.799-810, 1995.
- 5) 糟谷志帆, 井上亮: 階層的な隣接地域結合操作に基づく点事象の時空間集積検出の提案, 地理情報システム学会講演論文集, Vol.21, No.C-6-3, 2012.
- 6) Inoue, R., Kasuya, S., and Watanabe, T.: Spatio-temporal cluster detection of point events by hierarchical search of adjacent area unit combinations, *13th International Conference on Computers in Urban Planning and Urban Management (2-5 July 2013, Utrecht, Netherlands)*, p. 51, 2013.
- 7) 井上亮, 糟谷志帆: 複数の柔軟な形状を持つ点事象集積地域の検出法の開発, 土木計画学研究・概要集, Vol.49, 2014.
- 8) 井上亮, 塚原元英: 形状複雑度を制御可能なネットワーク上の点事象集積領域検出法の提案, GIS-理論と応用, Vol.24, No.1, pp.39–48, 2016.
- 9) Kulldorff, M.: A spatial scan statistic, *Communications in Statistics: Theory and Methods*, Vol.26, No.6, pp.148–196, 1997.
- 10) Shiode, S.: Street-level spatial scan statistic and stac for analyzing street crime concentrations, *Transactions in GIS*, Vol.15, No.3, pp.365–383, 2011.
- 11) Kulldorff, M., Huang, L., and Konty, K.: A scan statistic for continuous data based on the normal probability model, *International Journal of Health Geographics*, Vol.8, No.58, 2009.
- 12) 井上亮, 渡邊拓也: 不動産取引事例情報に基づく取引位置・価格水準の時空間集積の抽出, 地理情報システム学会講演論文集, Vol.20, 2011.

(Received October 1, 2021)

(Accepted *** **, **)

EXTRACTION OF MAJOR INTERSECTIONS BY ONLY LARGE SCALE VEHICLE TRAJECTORY DATA

Hengyi ZHONG, Wataru NAKANISHI, Shohei YASUDA, Takamasa IRYO

Most of researches of road network generation by vehicle trajectory data are focused on generating a complex network which can be used in car navigation. But there is no research on generating a network with major roads and intersections for analysis. This paper introduces a simple way to extract major intersections by using vehicle trajectory data only. We checked the distribution of trajectory data around intersections, and applied the spatial scan statistics with two kinds of distributions, the angle distribution of azimuths, and the spatial distribution of data with large turning angles. We extracted areas with large variance on angle distribution, as well as areas where data with large turning angles concentrate, as intersection areas, and then tried to build a network with extracted intersections. As a result, a possibility of generating a network for analysis by using large scale vehicle trajectory data is shown in this paper.