

1. はじめに

本稿の目的は、都市交通計画で用いられているモデルとデータについて発展の経緯を俯瞰した上で、移動体通信システムを用いた位置データの役割を再定義し、今後の研究の方向性を見通しのよいものにすることにある。今後の発展が見込まれる移動体通信システムによる位置データ取得を基本にしたプローブパーソン調査の今後の可能性と課題について著者なりの考察を加えたい。

2. 従来の計画手法の問題点

図-1に交通分析手法の変遷を整理する。1953年にデトロイトで行われた大都市圏交通計画調査 (DMATS) を契機としてさまざまなモデルが開発され、交通計画における科学的な分析が用いられるようになった。

初期の交通ネットワークの分析モデルは、統計力学と交通現象に類似性に着目した物理学からの転用が目立つ。統計力学や電気ネットワークの理論に対して交通現象を当てはめられた結果、その有効性が確認されたが、交通現象に特有の旅行者の確率的な行動現象に注目する必要があった。1950年代、NewellやSasakiらが交通流のモデル化を図られ¹⁾るとともに、物理学のアナロジーに着目した重力モデルやFrator法などが開発され、実務においても科学的なモデルが徐々に用いられ交通計画が行われることになった。

1967年の広島都市圏で行われたパーソントリップ調査調査を端緒として、定量的な分析に耐える一定のデータストックをもたらされた。パーソントリップデータを用いた定量的分析によって明らかになったのは、四段階推定法のモデルの再現性の低さであった。四段階推定法全体を一つのモデルとして考えた場合、論理的一貫性に欠けていることが指摘される。こうした問題は心理学系のモデルでも同様に指摘されている。ここでの論点は、サブモデルとしていくらか精緻なモデルであってもシステム全体として考えると「抜け」が多い点にある。「抜け」

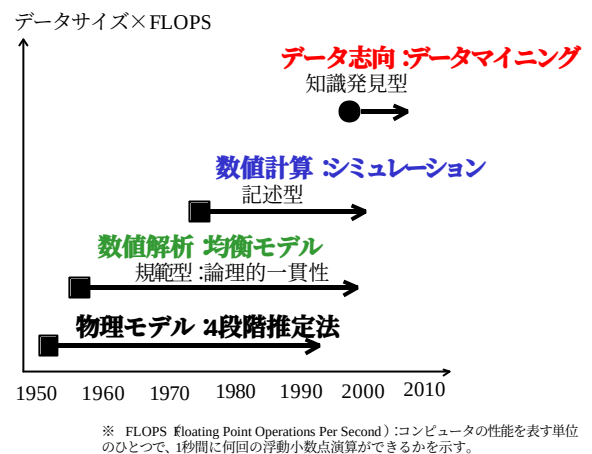


図-1 交通分析手法の変遷

を意味する。こうしたモデルで、交通現象を個人の意思決定から都市圏レベルの交通渋滞まで分析した場合、過大あるいは過小な推計結果をもたらす可能性がある。無限小、無限大を扱えないアドホックな数値計算のつなぎ合わせでは誤差は至る処で入って来るからである。

交通システムの様々な特性を見通した上でたてられる交通施策立案の作業において、複数のサブモデルの組み合わせについて系統的に吟味しないまま用いることは政策評価を簡単にする反面、全く当たらない予測結果をもたらすかもしれない。

システム全体の挙動がミクロな挙動の単純な積み上げにならないことは物理現象における繰り込み群や、システム制御の分野で既に議論されている^{注1)}。個々の現象を構成する粗格のスケールでおきている現象はある程度調整をされた上で、その前後のスケールで起きている現象に影響を及ぼす。異なるスケールの現象間のコヒーレントな関係は四段階推定法に限った話ではない。ミクロな意思決定一本でマクロレベルまで説明してしまう均衡配分にも同様な問題は存在する。サブモデル間で発生する状態変数のくりこみ方や、その集約の方法が重要となる。

このような問題は、道路整備のように長期的な施策の効果を大雑把に予測する場合よりも、短期的で詳細な施策による交通状態の変化を正確に予測したり、そ

*連絡先 : ☎790-8577 松山市文京町3 愛媛大学
TEL089-927-9830, Email hato@eng.ehime-u.ac.jp

の結果を用いてリアルタイムに交通の状態を制御する際に重要となる。移動に関するデータ革命が可能にしつつあるCRMや、それを活用したYield Managemnetなどがこうした問題に直接関連してくる。そして、こうした問題を取り扱うためには様々なスケールの系における観測データと、その上位-下位における同時観測データを必要となる。逆にいえば、こうしたデータがなければスケールの問題やモデルの一貫性が問われることはない。取得可能なデータの範囲にモデルのcapabilityが限定されるためである。

3. データの役割

配分計算における経路の取り扱いを考えてみよう。通常の確率的利用者均衡問題では、経路は中間変数として処理される。OD表とリンクコストパフォーマンス関数を入力変数として、リンク交通量を直接出力する。これは、パーソントリップ調査のようなアンケート調査では利用した経路を被験者にいちいち記入してもらい、経路のデータを得ることが困難であるためである。モデルの検証は経路交通量ではなく、リンクデータを用いて行われる。こうした計算アルゴリズムでは、経路に対して解の唯一性は保証されない。

また配分計算の過程で、ODペア毎に経路のラベリングを行うためには、k番目最短経路探索を適用しなければならない。この際、ラベリングした経路を、その都度計算機のメモリーに格納する必要がある。膨大な記憶容量が必要とされ、現実的なネットワークで経路ベースの配分を行うことは現実的ではない。これらの理由により、実際に経路そのものを明示的に取り扱ったモデルが使われる例は少ない^{注2)}。

パーソントリップ調査では、5%程度の抽出率で1km²程度のゾーン単位にトリップを集約した上で、路側の観測交通量を基にモデルの検証が行われてきた。交通現象をモデル化する際、無意識のうちにこうしたデータを前提としてきた。こうした調査手法では現実的な経路データを収集することは殆ど不可能である。

一方、近年の交通分野におけるデータ革命では、携帯電話やPHSによる旅行者の移動データの取得が可能になっている。こうしたシステムを用いたプローブパーソン調査では、シームレスに人の移動-活動データを識別することができる(羽藤 朝倉,2000)¹¹⁾。セル単位で設定されている通信基地局密度に応じて500m～

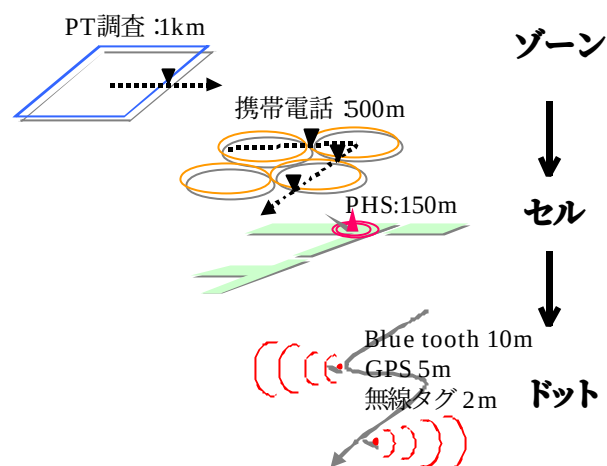


図-2 交通調査におけるデータ取得スケールの変化

50m程度の位置精度が得られる。

PDAのBlueetooth機能やGPS、無線タグなどを用いることで、10mから2m程度と、ドットスケールの精度で旅行者の位置計測が可能となっている。こうした高精度な旅行者の移動データは、通信ネットワーク上で相互接続されることで、膨大かつ正確なデータベースとして構築されようとしている。

このようにして生成した移動に関する膨大なデータフローがモデルにどのような影響を与えるかを考えてみたい。従来のアンケート等によるデータ観測アプローチを、ゾーンや断面に着目して旅行者の動きを把握しモデル化しようとしたオイラー的アプローチとするなら、旅行者そのものの動きをドットで縦断的に計測するプローブパーソン調査は、ラグランジュ的なアプローチとして定義できる。

ラグランジュ的な移動データが移動体通信システムなどを用いたプローブパーソン調査によって観測可能になり、計算機能力が向上すれば、従来中間変数として取り扱われてきた経路交通量を明示的に取り扱ったモデルのハンドリングが可能となる。また、複数日のデータを用いたクロスバリデーションチェックの導入により、モデルの再現性の確認方法は変化するであろう。day-to-dayの交通行動やそれに関連したネットワークの状態の揺らぎが観測可能になれば、本質的に交通現象への理解が進む。設定しているモデルフレームそのものが見直されることも期待できよう。

4. データオリエンテッドアプローチ

データを用いてモデルの前提条件や再現性を検証

しようとする従来のアプローチに対して、交通政策に有効な戦略を多量のデータから直接掘り起こすことは難しいだろうか。本節では、データ志向の新たな分析アプローチとして、データマイニング(Data Mining)を考える。

データマイニングのよく知られた適用例として、米国最大の小売業であるウォルマートの「紙おむつと缶ビール」の話がある^{注3)}。POS(Point of Sales)システムにより瞬時に収集される膨大な購買データに対してデータマイニングの代表的な手法である相関ルール分析を行った結果、「金曜日の夜、紙おむつを買う客は缶ビールをよく買っている。」というルールが発見された。ウォルマートは早速、紙おむつ売り場の傍に缶ビールを置いてみたところ、売上が倍増したというものである。

このエピソードは、分析者が直感として得ることが難しい消費者の行動の因果関係を多量のデータから直接引き出すことが可能であることを示している。現在では、在庫管理・新商品企画・証券評価・株価評価・医療診断など、様々なビジネスデータ解析にすでに利用されており、めざましい実績をあげている。

従来の分析アプローチがモデルの一貫性や再現性を重視するあまり、政策評価の柔軟性に欠けるのに対して、データマイニングは多量のデータを背景に、交通政策をマーケティング的に捕らえ、なんからの有効な関係性をデータから見つけ出すことに注力している。

データサイエンスに基づいたこうした方法は従来の「データサンプリング」→「解析」→「モデリング」→「検証（バリデーション）」→「現実問題の解決」というアプローチと大きく変えることはない。ただ違うのは、多くの関係性をデータ自らに語らせようとする点である。そのためには、正確なデータを蓄積運用していくと共に、蓄積された多くのプローブパーソントリップデータを使うことで、なんらかの意味のある交通現象の関係性を掘り出すための技術が必要となる。データの効率的な蓄積方法とこうした「宝」を掘り出すための手法論がデータマイニングには求められる。

5. 交通データマイニング

現在、交通工学の分野でも、プローブビークルやビーコン、感知機、電車の自動改札、ETC、GIS-tなどにより膨大な移動データの収集と蓄積が可能となっている。こうしたデータを基にした交通データマイニングの方法を示す。

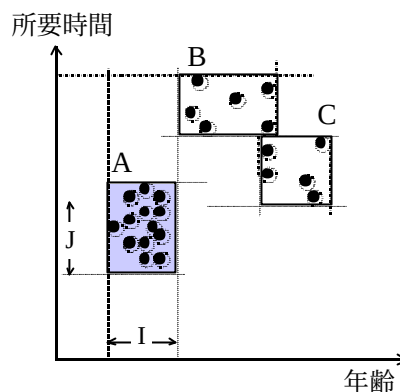


図-3 公共交通機関の1次元相関ルール

公共交通の利用者の特徴分析を考える。交通データマイニングでは、年齢と所要時間を使って公共交通機関の利用 を特徴づける結合ルールを見つけ出す。

(年齢 in I) AND (所要時間 in J) => (公共交通機関)

条件部の範囲を年齢と所要時間の張る2次元平面で図示すると、図-3の矩形領域になっている。このような矩形領域は、公共交通機関を利用する顧客のいる範囲を大まかに教えてくれる。また、領域に入る顧客数の最低値をユーザーが与えると、公共交通機関 を利用する顧客の確信度の大きい領域Aを見つけることが興味の対象になる。1次元の場合に導入した最適確信度結合ルールの発見問題は2次元以上の場合にも自然に拡張できる。このようにして、公共交通機関を利用する顧客を高い確信度で特徴付けることができれば、公共交通を運用している経営者に対して有効なマーケティング戦略を立案することが可能となる。

次にプローブパーソン調査から得られる膨大な位置データを利用した空間データマイニングについて説明する。ラグランジュ的な移動データとGISの空間情報とマッチングさせることで、頻出する都市の施設の連続的な回遊パターンを抽出し、特定の商業施設と他の施設との相互関係を分析することが可能である。

図-4は松山市中心市街地において移動体通信システムを用いて収集された位置データをプロットしたものである。商店街Aでは回遊する複数施設の組み合わせパターンが多いけれど、商店街Bでは有意なパターンを見出せない、あるいは、ある百貨店に行った人はAという商店街までは行 けれど、Bという商店街までは行かないとか、Aという商店街の服飾店に行った人は半径

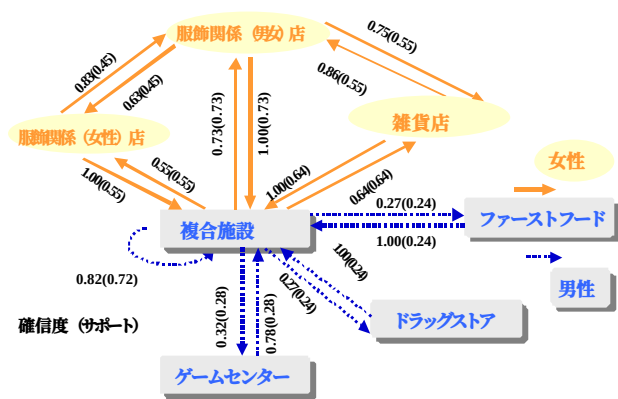
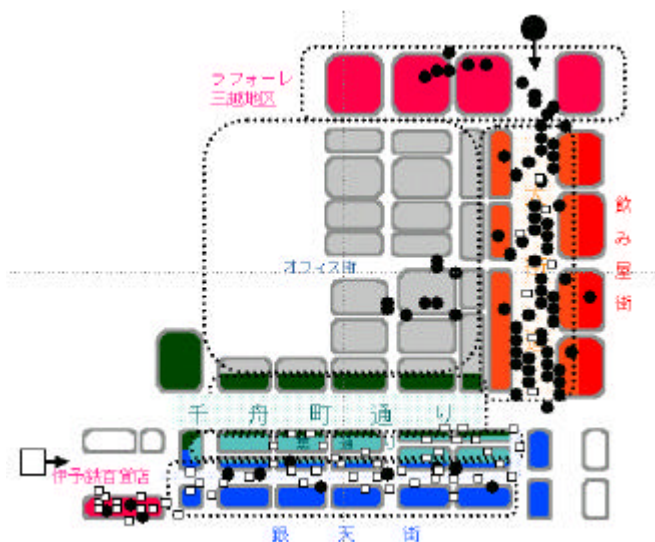


図-4 都市内回遊のデータマイニング

300m以内の雑貨屋に行くとかいった頻出パターンが抽出できる。上の例では女性と男性で抽出される行動文脈のパターンが異なっていることがわかる。このような行動文脈のパターン分析から、回遊性を高めるための中心市街地の施設の再配置戦略の立案が可能である^{4,5)}。

投資予測にデータマイニングの技術が用いられる現状を踏まえると、交通需要予測においてもデータマイニングは有効であろう。短期の需要予測は、ある時点の交通量をその日の予測時点までの観測交通量を元に予測する。この場合、交通需要の変動の非線形な関係性を膨大なデータベースから抽出することが考えられる。一方GWの交通量をGW前に予測するような長期の需要予測においても、多数の要因の非線形な関係が存在するため、データマイニングのような手法が適していると考えられる。

6. プローブパーソン調査は宝の山か？

データマイニングはデータの背後に潜む構造なり現

象の本質を掘り出すための手法といえる。しかし、掘り出すべきデータの山に「宝」が埋まっているかどうかは、プローブパーソントリップ調査によって得られたデータそのものをどのように整備していくかに依存するところ大きい。以下に、データベースを構築していくための課題をまとめる。

交通データマイニングでは多量かつ正確なデータベースを必要とする。交通データのような追記型データベースでは、レコード数も属性数も増えて続けることになる。並列処理やインクリメンタル処理はギガからテラに及ぶ規模のデータを処理するために欠かせない。従来のような市販の表計算ソフトによる集計解析の手法論ではこうしたデータの取り扱い不可能である。

GPSや携帯電話やセンサー/検知器などのパーベシブコンピューティング環境下で得られるデータはそれだけでは分析可能なデータにはならない。これらを共有可能なデータにするためには各デバイスにおけるインターフェースの正規化が重要である。たとえば車速データは通常パルス(Hz)として計測され内部処理されるが、そのまま公開したのでは、速度と判別することは難しい。車速パルスを交通データとして共有化していくためには外部に対してパルスでなく、速度として公開しなければならない。システム全体におけるデータの利用法をある程度想定し、その上でデータ構造の標準化と公開を行なうことが必要になる。また分析目的にあった知識をデータから抽出するためには、土地利用などの空間データや経済データを同時に整備し、XMLなどによって相互利用を図っていく必要性が高い¹²⁾。

プライバシーとセキュリティも大きな問題である。100万人の規模の都市で、個人の移動データ（識別番号、時間、緯度、経度）を一分間隔で一年間記録すれば、100TB以上の大規模な個人データベースが構築されることになる。セブテンバーイレブン以降、本土防衛の観点から大規模な個人情報システムの導入を期待する向きもあるが、プライバシーの観点から一挙監視型社会への危惧は根強く、アノニマスの確保のための技術が必要である。

プローブパーソン調査を通じて得られるデータは基本的に個人情報であり、個人情報保護の観点から処理されることが望ましい。重要なものではデータの利用法を調査者自身が個人情報ガイドラインを作成した上で定義した上で、その利用法について被験者との間に合意

契約を結ぶことである。この際、個人情報の管理者を設定した上で、SSLなどの暗号技術を用いて調査を実施し、データ管理をすることが必要である。

7. モデルとデータ

移動の計測技術も計算機技術も、その進歩は留まるどころを知らない。しかし、そもそも計算する対象がはっきりしないときには計算機でプログラムを組むという作業が出来る筈もない。そうした場合には紙と鉛筆で計算できる理論研究が優れているとされてきた。

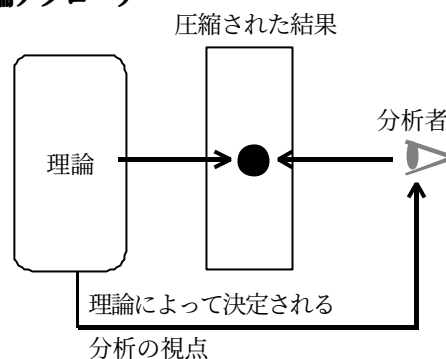
しかし、データマイニングでは、プローブパーソン調査によってデータベースに膨大にストックされたデータをもとに、なんらかの知識を抽出するものである。こうした手法論は分析者すらさえも気づかない因果関係を自動的に見つけ出すことに主眼がおかれている。双方のアプローチは構造的に異なっており、前者のアプローチが近代的であるとするなら、後者のアプローチはデータ革命を下敷きにした新たな手法論であり、ポストモダンなアプローチとして位置付けられよう。

ここで従来のアプローチとポストモダンなアプローチの構造の違いについて考えてみたい。理論アプローチでは、モデル全体が、大きな一つの仮説を通して構成されており、得られる結果はその仮説に限定され、圧縮された結果を得ることができる。このため、その分析者が用いるモデルそのものが分析の視点を決めることになる。一方、シミュレーションアプローチでは、モデルそのものは複数のサブモデルから構成され、得られる結果は組み合わせられるサブモデルの構成や多数のパラメータチューニングの設定によって異なる。分析者によって結果に対する柔軟な対応が求められることになる。

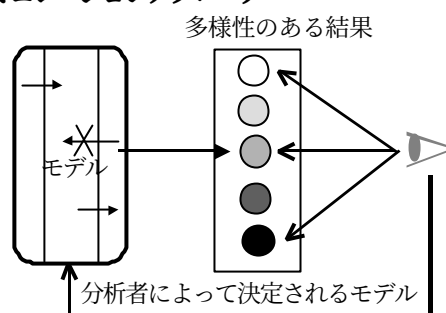
こうした分析アプローチでは、モデルが結果の範囲を決めるにしろ、分析者が使うモデルを選択するにしろ、モデルという、「交通現象に対するものの見方」というべき型が存在していた。分析者と分析対象の関係は投射構造を成しており、その中では、交通現象の大きな構造を明らかにすることに主眼が置かれていたといえる。

一方、データマイニングには、現象を見るための型は存在しない。この点においてデータマイニング以前のア

理論アプローチ



シミュレーションアプローチ



データマイニングアプローチ

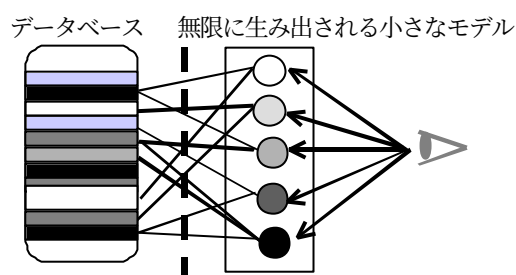


図-5 分析アプローチの分類

アプローチとデータマイニングの間には大きな違いがある。データマイニングのアプローチにおいてはデータベースさえ整備されていれば、無限の小さなモデルが自動的に生成される。分析者はこうして得られるモデルを「知識」として活用し、交通政策決定に役立てる。多量のデータをざっと見渡し、目的達成のために有効な知識を効率的に抽出するためには、知識体系と評価基準を定義する必要がある。このための技術はKDD (Knowledge Discovery and Data mining)と呼ばれる。

データマイニング以前のアプローチが投影型であるのに対して、こうしたポストモダン的な分析アプローチは、データや知識をある設定の下で分析者が読み込む「読み込み型」として定義できる。読み込み型はデータオリエンテッドであるため、多量でかつ正確なデータベース

さえ準備されていればプッシュ型の交通戦略を検討する場合に有効である。しかし、こうしたアプローチは何もデータマイニングに限った話ではない。ドットレベルの豊富なデータに基づいて詳細な解析が行われることで、個々の交通現象についてみれば様々な知識が生み出されていくことが予想される。しかし、その反面、分析者は現象全体を理解するための大きなモデルを失うことが考えられる^{注4)}。

果たして、どちらのアプローチがより優れているのだろうか？理論モデルはよく絵画に例えられる。デッサンの練習は困難であり、所定の作法を身に付けた上で、画家は新しい画法を求めて絵を描くけれど、平面のキャンバスには限りがある。シミュレーションのようなコンピュータ・グラフィックスが出てきてしまえば動画で様々な画像を描き出すことが可能となる。

データマイニングは、型を持たず、ハイパーリンクをたどることで様々なコンテンツを用意に手に入れることができる。こうしたアプローチは、インターネット的ともいえるし、精密さからいえば集積されていくデータベースは航空写真のようなものかもしれない。いずれにせよ、膨大かつ正確なデータベースを元にして検証されるモデルと、機械仕掛けで導かれる様々な知識は交通工学にとってきわめて有益となる可能性を秘めている。

絵画と航空写真のどちらが美しいと思うだろうか。航空写真は精密である。しかし、それだけではどこをみればいいかわからないし、長期的にどこに向かえばいいかもわからない(どちらへ向かうかを決める役には立つけれど)。「絵画」である理論の美しさには何をもってしても叶わないように著者は思う。平面のキャンバスに苦勞した末、凝縮された結果は対象の本質を見事に引き出す。しかし一方で、絵画の描き方や本質を抽出し、美しさを感じるための努力を描き手と見る側が怠っては、「絵画」が「航空写真」にとって代わられることも、いつの日かあるのかもしれない。

付記：本稿は雑誌交通工学に掲載された「ポストモダン交通工学」に加筆修正を行ったものである。

注1) この議論は2003年3月に東北大学で開催された研究ワークショップ「モデルとデータ？ミクロとマクロの関わり合い」において赤松氏によって示された論点である。経済学の主張するモデルの論理的な一貫性と心理学の主張する個別現象記述力の高さの双方に対する問題提起である。

注2) 経路ベースの配分アルゴリズムについてはDamberg et

al.(1996)がModified DSDを開発している。しかし逐次的な端点記憶に膨大なメモリーが必要とされる。

注3) ウォルマートはPOSシステムをとりいれデータマイニングを活用し、売上向上にもっとも成功した企業として知られる。一方、1990年代初めにウォルマートのライバルであったKマートはPOSシステムの導入に長らく躊躇した結果、2002年1月に破産宣告している。高速料金の設定やバス路線設置、交通情報が完全に自由化にされれば、航空業界が1970年代Yield Management理論を導入し徹底的な顧客管理を行ったように、道路利用者のデータマイニングに基づく顧客管理が求められよう。

注4) 様々なデータが収集された結果として、大きなモデルに対する信頼が揺らぎ喪失し、分析者がデータオリエンテッドに小さな知識サブセットを利用するという考えは、複雑系などの様々な科学分野で多く見られる。こうした状況は、近代以前の大きな権威が近代以降失墜して、全く新たな文化を生み出していった時代背景をポストモダンと定義する考えに近い。

参考文献

- 1) Newell, G. F.(2002)Memories on highway traffic flow theory in the 1950s, Operations Research Vol.50, No.1, pp.173-178, 2002.
- 2) 羽藤英二, 朝倉康夫(2000). 時空間アクティビティデータ収集のための移動体通信システムの有効性に関する基礎的研究. 交通工学, Vol.35, No.4.
- 3) 交通工学研究会 交通シミュレーション自主研究委員会(桑原雅夫委員長) 交通シミュレーションクリアハウス. <http://www.jste.or.jp/sim/activity/>
- 4) 羽藤英二 他 (2002)都市回遊行動に関する空間データマイニング, 土木計画学講演CDROM
- 5) 羽藤英二(2003)ドットデータを基本にした空間データマイニング, 土木学会年次講演集CDROM 掲載予定)