

プローブカーによる交通データ測定の課題について

— 行動履歴データに対するプライバシー保護の観点から —

南山大学 情報理工学部 システム創成工学科
河野浩之 (kawano@nanzan-u.ac.jp)

1. はじめに

環境，エネルギー，景観などの問題に道路交通は深く関わっており，加えて，国際化や高齢化に関わる道路交通の諸課題の解決において，交通データを正確に把握する技術は重要である．世界的には 1950 年代以後，日本では 1960 年代以後，様々な理論や技術に基づく交通データ測定が行われ，道路網整備，交通管制，交通情報提供等に用いられてきた[4]．例えば，安全・快適な交通管理を行うため，交差点の幾何構造や交通量による交通信号制御，道路への流入・流出を制限するランプ制御，利用者に対する所要時間や渋滞情報や事故情報を VICS (Vehicle Information and Communication System)での情報提供による経路選択などである．

著者らは，多様な交通データを蓄積する交通データウェアハウスの構築に関して，データベース，データマイニングの観点から研究を行っている[6,7]．PHS (Personal Handy-phone System) を用いた行動履歴データの取得や，GPS (Global Positioning System) を備えたプローブカーやフローティングカーと呼ばれる車両に搭載された各種センサーからリアルタイム，もしくは，オフラインで収集されるデータの利活用を試みている．交通データウェアハウスには，出発地・目的地や経路といった位置情報のみならず，車種データや気象データも蓄積される．過去の研究においては，パトロールカーにより記録されるデータには，巡回監視や障害物対応のログに加えて，個人情報を含む事故処理対応のようなログを含むため，プライバシー保護に細心の注意を払う必要があった．

本稿では，交通データウェアハウスに格納されている超音波感知器などと，プローブカーとしてのパトロールカーにより収集された交通データを融合するデータ処理について紹介する．また，位置情報を含む行動履歴データに関わるプライバシー保護を扱うため，プライバシー保護と関連するデータマイニングやデータ同定の研究を調査する．以下，2 章では位置情報を伴う行動履

歴データの利活用の状況とプライバシー保護について、3章は交通データ測定手法と精度について、4章ではプローブカーを用いた交通データ測定の研究動向について紹介する。そして、5章をむすびとする。

2. 行動履歴データの利活用とプライバシー保護

2.1節では、位置情報を伴う行動履歴データ利活用に関する現状と課題について紹介する。2.2節では、行動履歴データ処理に関わるプライバシー保護技術について、データマイニングとデータ同定の側面から整理する。

2.1 行動履歴データ活用の現状

「2001年度インターネット ITS 名古屋実証実験」でプローブカー利用による交通データ取得実験が行われる等(<https://www.mlit.go.jp/pri/houkoku/gaiyou/pdf/its.pdf>)、様々な観点から位置情報を含む行動履歴データの利活用が研究されてきた。もちろん、位置情報はプライバシーと深く関わるため、利活用において慎重さが必要であることは、俗称「Nシステム」として知られる自動車ナンバー自動読取装置によるデータ収集の議論から明らかである。

にもかかわらず、位置情報を伴う行動履歴データを収集蓄積する情報システムは急増しており、長期的に蓄積されたビッグデータとしての利活用について盛んに議論が行われている。例えば、自動車会社は、カーナビゲーションシステムにスマートフォンを連携することで「通れた道マップ」(<http://map.g-book.com/>)などのサービスを提供している。また、位置情報を収集する携帯電話会社により位置データ取得ビジネスが活発化し、シンポジウム[11]では、株式会社NTTドコモから、防災計画、交通計画など「モバイル空間統計」サービス事業について、KDDI株式会社からは、観光振興拠点をエリア別分析することで観光動態調査について、株式会社Agoopからは、ソフトバンクのモバイル事業関連分析で月間10億件の通信ログによる流動人口や商圈分析を行っているとの報告があった。いずれも、プライバシー保護に関して、運用データに対する非識別化処理の後に集計処理・秘匿処理を行うとしている。

しかし、2013年に、ICカード乗車券システム[5]により収集されたSUICAデータの利活用に関して、個人情報に関する利用者許諾の形態や、蓄積処理されたデータの第三者利用にかかる課題が報道されたことは記憶に新しい。2013年夏のプログラミング・シンポジウム「ビューティフルデータ」(<http://atnd.org/event/spro2013>, 2013年8月15日開催)では、履歴データ

の利用を行う「委託と第三者提供」にかかる問題点が、招待講演「パーソナルデータ保護法制に向けた最近の動向」（高木浩光）で報告された。

なお、わが国におけるパーソナルデータ収集に関する制度の動向は、「パーソナルデータの利活用に関する制度見直し」が参考になる(<http://www.kantei.go.jp/jp/singi/it2/pd/dai5/siryoku2-1.pdf>)。関連して、氏名、取得目的、取得方法、利用目的、通知・公表や同意取得方法、第三者提供、ポリシー変更手続き等に関するプライバシーポリシーの作成(http://www.soumu.go.jp/menu_news/s-news/01kiban08_02000087.html)や、国土交通地理空間情報プラットフォーム(<http://www.spat.nilim.go.jp/home/>)などがある。また、国境を越えてデータ流通が生じることを考えれば、国際的なプライバシー保護制度の整備を必要としており、2014年4月欧州議会本会議で採択の見込みである「EUデータ保護規則案」などにも注意を要する。

2.2 プライバシー保護データマイニングとデータ同定技術

適切なプライバシー保護制度の下で収集・蓄積されたデータベースに対してプライバシー保護を行ったデータであっても、複数データベースに対するデータマイニングやデータ同定の技術を用いて処理する場合のデータや抽出ルールに関しては、再度、プライバシー保護を考えねばならない。特に、2003年5月に個人情報保護法が制定されて以後、この種の位置情報を含む履歴データの収集技術や蓄積技術、さらに、データマイニング技術が飛躍的な進歩を遂げており、プライバシー保護を必要とするデータ処理技術を正しく把握する必要性が高い。すなわち、個々のデータベースにおいてプライバシー保護がなされていても、複数データベースの融合に対してプライバシー保護されずとは限らない。以下、プライバシーの観点からデータ処理におけるデータマイニングとデータ同定に関して簡単に整理する。

プライバシー保護データマイニング(Privacy Preservation Data Mining)の研究は、垂直分割モデルと水平分割モデルが提案されている[13]。また、各種統計関数で処理する場合のデータのランダム化処理だけでなく、「決定木、ベイジモデル、回帰モデル、相関ルール、クラスタリング、外れ値など」主要なデータマイニング手法により求められるルールを対象とする。そこで、プライバシー保護のために、識別属性を削除する匿名化、暗号化されたデータに対するアルゴリズム、データベース中の属性削除(Suppression)や属性値の仮名化(Pseudonymization)や一般化(Generalization)、かく乱(Perturbation)、さらに、 k 人未満に絞り込まれない k -匿名性などの手法が研究されてきた。もっとも、統計データベースにおける機密性に関する問合せ回数や構造に関わる制約の議論は1970年代から行われており、蓄積されたデータに対するプライバシー保護は非常に困難な問題である[12]。

他方、データに欠落があるレコードの統合やリンク解析 (Data Integration and Link Analysis) を対象とするデータ同定の研究[2]があり、「匿名化, 暗号化, ランダム化」等と相反する側面をもつ。加えて、データ同定は、データマイニングと深く関係し、データの前処理における各種セグメンテーション、インデックス技術としての Q-Gram, 属性レコード比較のための各種距離関数や TFIDF などが用いられる。また、暗号化技術に基づくプライバシー保護マッチング (Privacy-Preserving Matching) の研究がされるなど、完全一致以外に近似マッチングやスケーラビリティなど、幅広い研究がなされている。

したがって、データマイニング研究に関連するプライバシー保護とデータ同定という異なる側面に注意しながら、プライバシー保護データマイニング技術の検討を要する。なお、データマイニングに関する国際会議の一つである IEEE ICDM における国際ワークショップ (DPADM, International Workshop on Discrimination and Privacy-Aware Data Mining) などにおいて、プライバシー保護が技術面だけではなく社会制度に及ぼす影響を含めた議論がなされており参考になる。

3章 交通データ測定の課題

3.1 節では、交通データ測定手法について、アンケートによるパーソントリップ調査、ならびに、感知器などを用いたデータ収集手法について整理する。3.2 節では、交通データ測定機器の特性についてまとめる。

3.1 交通データ測定手法

1950 年代以後蓄積されている交通データとして「パーソントリップ調査」(Person Trip Survey) がある[4]。位置情報の精度や付随的な行動データの z 属性は、調査主体となる国や組織により異なるが、経年比較可能な継続性のあるデータ収集が長期的になされてきた。日本では 1966 年の試行調査以後、住民基本台帳に基づくアンケートベースの家庭訪問調査がなされており、全国 70 都市程度、1 都市あたり 500 世帯の規模で実施されてきた。調査は、特定の代表日に 10 年から 20 年周期で 1 回実施されているため、特定都市の年次的変化を把握することは難しい。そこで、補完調査として、乗降客調査などが行われてきた。

また、アンケートによるパーソントリップ調査以外に、情報処理技術の進歩に伴い、区間観測や地点観測を行う様々な交通データ測定技術が導入されてきた。例えば、道路上に設置されている車両感知器、テレビカメラ、車両番号読取装置 (AVI)、突発事象検出カメラなどの交通データ測定機器である。

渋滞検知や所要時間算出において主要な役割を担う超音波車両感知器は、高速道路上の 500m 程度の区間に一つ設置される。構造がシンプルであり、比較的安価であることから、一般街路にも数多く設置されている。また、路面に埋設するループ式車両感知器もあるが、メンテナンスが困難であることから導入数は多くない。ただし、超音波車両感知器とループ式車両感知器のいずれも、異なる地点を通過する車両の同一性の判定が正確にできないため、精度の高い OD 表作成には適していない。

他方、テレビカメラを用いた AVI や突発事象検出カメラは、ナンバープレートなどの自動認識により同一車であることが判定で可能であるが、多くは 100m 程度の区間を監視する目的で設置する装置であるため、道路網全体の交通状況を測定するには、価格やメンテナンスの課題が大きい。

その他、料金収受が主要目的である ETC (Electronic Toll Collection System) の導入が、2001 年以後進んでいる。双方向通信が可能な DSRC (Dedicated Short Range Communications) を用いることから、情報提供にも利用される。交通データの測定精度向上に役立つ技術開発や実証実験などは数多く行われており、例えば、路車間通信や電子ナンバープレートによる車両識別などをあげることができる。

さらに、文献[8]のようにプローブカーによるデータ収集と VICS データを融合した旅行時間推定の研究から、各種データの推定精度が変化している。今後、24 時間 365 日モバイル機器による収集データをパーソントリップ調査に用いることによるデータ品質向上に期待される。但し、大規模な行動履歴データ利活用に対する社会的コンセンサスの形成が必要不可欠である。

3.2 交通データ測定機器の特性

道路交通流に関する特性は、交通量 q 、速度 v 、交通密度 k といった交通データの測定により表される[4]。交通量 q は、一日あたりや時間あたりの単位時間交通量を表す。速度 v は、地点速度ならびに旅行速度（区間速度）や、時間平均速度や空間平均速度により表される。交通密度 $k (= n/L [台/km])$ は、長さ L の観測区間に n 台の車両が存在していることを表す。また、一定区間の道路長に含まれる全ての車両の合計長である空間占有率の実測が困難であることから、時間占有率が用いられることが多い。そこで、前節に述べた超音波車両感知器、AVI、ETC、プローブカーなどの測定手段を複合的に用いることで、道路交通流に関する特性を把握する交通データ測定の精度向上が期待されている。

道路交通状況を正確に把握するには、データ測定精度の向上が必要である。例えば、交通量が最大となる臨界点の存在は、存在速度-密度曲線 (speed-density curve) または K-V 曲線の関係から求まる。その他、交通量-密度曲線

(flow-density curve), 交通量-速度曲線 (flow-speed curve) または Q-K 曲線, Q-V 曲線などが用いられる. また, 精度の高い OD 表(Origin-Destination table)の作成は, 道路ネットワークの経路選択を伴う交通量配分に重要な役割を果たしている. なお, マルコフ過程による数学モデルを用いた交通量配分, 確率統計的モデル, 待ち行列モデル, 流体モデル, 追従モデルなどによる車両走行に関する研究等が行われてきた.

表 1 に, 超音波車両感知器, AVI, ETC, プローブカーによる各種測定機器の特性と, 各種測定データのおよその特徴を示す. 測定精度は, 設置環境 (街路, 高速道路等) や測定条件 (渋滞時, 昼夜間等) に大きく依存している. なお, センサーを車載するプローブカーは, 車両周辺の交通データを連続的に収集することが可能であり, スマートフォン利用によって通信コストを含めて優位であると考えられる.

表 1 各種測定機器の特長と測定データ

測定	超音波	画像	ETC, DSRC	プローブカー
交通量	正確 (誤差)	正確 (誤差)	正確	不確か
交通密度	30 秒間隔 (誤差)	正確 (誤差)	—	不確か
速度	30 秒間隔 (誤差)	正確 (誤差)	—	正確
車種判別	車両高・車両長	およそ正確	正確	正確
OD 表作成	困難 (誤差大)	およそ正確	かなり正確	積載車は正確
設置位置	500m 程度間隔等	120m 間隔等	出入口等	車載
測定位置	機器直下	撮影領域	機器直下	車両位置
画像	無し	有り	無し	車両周辺
コスト	安価	非常に高価	高価	比較的安価

4章 プローブカーによる交通データ収集

4.1 節で、プローブカーによる交通データ収集活用の研究の一例を示す。4.2 節で、プローブカーに関わる著者らの研究を紹介し、プローブカーによる交通データ測定精度向上の課題に関する考察を述べる。

4.1 プローブカーによる交通データ収集と活用の動向

交通データの精度向上においてプローブカーによるデータ収集が期待されており、既存の超音波車両感知器、AVI、ETCなどと融合する調査研究がなされている。例えば、桑原雅夫らは、始端・終端の車両感知器データと5台のプローブ車両データによる交通状況の把握を試みている。また、土生恭祐らと共に「プローブ情報と既存センシング情報を融合した車両軌跡の予測」(http://www.plan.civil.tohoku.ac.jp/kuwahara/research/Habu_track.pdf)の研究を行っている。なお、測定対象区間の途中に出入り交通がある場合にも測定精度の低下が生じるなどの課題がある。

その他、タクシープローブカーデータを収集し、旅行時間予測などを行う研究が複数有る[1,3]。中国大連市バス走行データに基づく適応回帰モデルを用いたバス到着時間予測の研究もある[14]。ただし、タクシーやバス等の営業車両は、客の乗降に伴う一時停止や客待ちの停車などが生じるため、一般街路の車両走行状況を正確把握することは難しい。

4.2 プローブカーデータに基づく速度測定と費用便益

阪神高速道路を2012年6月13日から7月14日の約1ヶ月間走行した複数のパトロールカーにより取得された、無線ID、年月日時分秒、十六方位、速度、世界測地系緯度経度等の値が記録されたプローブカーデータの提供を受けた。全ファイルサイズは2.14GB、走行ログ795,575レコードの内、取得成功件数が609,007レコード、エラーレコード数が186,568である。そこで、GPSベースのプローブカーの速度データと、超音波式車両感知器から得られたデータが格納されている交通データウェアハウスの両者に基づいて、走行区間全体の速度変化に基づくエネルギーコストについて議論してきた[6]。

なお、1号環状線、3号神戸線、4湾岸線、7号北神戸線、8号京都線、11号池田線、12号守口線等での走行ログが記録されているが、時間帯ならびにパトロールカー巡回頻度の偏りが大きく、十分なレコード数が記録されていないケースも多い。そこで、比較的レコード数の多い(158,341レコード、走行距離39.6km)3号神戸線を中心に、著者の研究指導のもと、竹内は阪神高速道路走行車両全体の最適速度を求め、費用便益の観点から燃費改善効果を中心に評価した[10]。論文では、運転支援や自動運転などの技術により、加減

速を最小限に抑えた安定走行を実現することが出来れば、阪神高速道路神戸線全体で1日あたり約278万円、1ヶ月で約8631万円、阪神高速道路全体では、1日約1952万円、1ヶ月約6億円のエネルギーコスト削減の余地があると見積もっている。

ところで、超音波感知器による交通データ測定は、長期にわたる技術的改善により交通量、交通密度、速度などの精度は向上しているとは言え、数百メートル区間の30秒間隔の交通データであり、大きな変動を捉えるには適しているが、OD表作成に必要なデータ測定は難しい。他方、パトロールカーをプローブカーとするデータ収集は、連続区間を実時間で交通データ測定ができ、プライバシー保護に関する課題も少ない。しかし、全体交通量に対してパトロールカーの密度は低く収集データ量も少ないため、既存の交通測定機器で収集したデータと融合できる区間は、非常に短時間である。

よって、パトロールカーのプローブデータに加えて、一般車両を含むETCデータや、その他のプローブカーから得られる交通データとの融合が不可欠である。そのため、複数プロセスから生成されるデータに対するプライバシー保護技術や、時系列データに対するアラインメント処理技術の検討を必要としている。

5章 むすび

本稿では、複数データベースから構成される交通データウェアハウスに、プローブカーにより収集される位置情報を伴う行動履歴データを格納し、より高度なデータ処理を行うために、データマイニングとデータ同定に関する研究動向をプライバシー保護の観点から調査した。今後、この種の行動履歴データ取得がより容易に安価になるにつれ、複数の交通データ測定技術を融合させることで、交通データの精度向上に結びつくと考えている。

今後、正確な交通行動が容易に把握できるにつれ、現在行われている系統制御や広域制御の交通信号制御以外に、動的な料金賦課による交通制御（ロードプライシング）導入も円滑に進むと考えられる。さらに、運転支援技術導入の本格化や、将来的な自動運転システムの実現は、交通制御や情報提供に大きな影響を与えると予想される。

謝辞

本研究は、2013年度南山大学パッへ研究奨励金 I-A-2「複数プロセスデータに対するアラインメントとプライバシー保護データマイニング」の支援を受けた。また、阪神高速道路株式会社から、パトロールカーによるプローブカーデータ提供を頂いたことに深謝する。

参考文献

- [1] E. Brockfeld, P. Wagne, B. Passfeld, “Validating travel times calculated on the basis of taxi Floating Car Data with test drives,” Proc. of 14th World Congress on ITS, pp.1-13, 2007.
<http://trid.trb.org/view.aspx?id=916158> (accessed 2014-02-28)
- [2] P. Christen: Data Matching, Springer, 2012.
- [3] T. Hunter, R. Herring, P. Abbeel, A. Bayen, “Path and travel time inference from GPS probe vehicle data,” Proc. of the NIPS, Vancouver, Canada, pp.1-8, 2009.
<http://bayen.eecs.berkeley.edu/sites/default/files/conferences/nips09.pdf>, (accessed 2014-02-28)
- [4] 飯田恭敬他: 交通工学, オーム社, 1992.
- [5] 伊藤裕一, 遠藤雅一, 宮武学 “スマートな乗車を支える IC カード乗車券システム,” 日立評論, 2010 年 8 月号, pp.54-57, 2010.
http://digital.hitachihyoron.com/pdf/2010/08/2010_08_08.pdf, (accessed 2014-02-28)
- [6] H. Kawano, “Analysis of optimal vehicle speed: from supersonic wave devices to GPS transceivers,” Proc. of 25th European Conference on Operational Research (EURO 2012) , 2012.
- [7] 河野浩之, “クラウドを活用した道路維持管理の可能性～阪神・速道路の取り組みから～,” 「これからの道路保全を考える」セミナー (講演資料), 名古屋通信会館, 2013 年 8 月 2 日, 2013.
- [8] 王立暁, 姜美蘭, 山本俊行, 森川高行, “プローブカーデータと VICS データの融合による旅行時間推定に関する研究,” 土木計画学研究・論文集, Vol.23, pp.1011-1018, 2006.
http://www.jsce.or.jp/library/open/proc/maglist2/00039/200506_no31/pdf/205.pdf, (accessed 2014-02-28)
- [9] 鈴木尋善, 欧米の協調 ITS システムと自動運転の最新動向, JARI Research Journal, No.20130901, pp.1-6, 2013.
- [10] 竹内秀樹, “プローブカーデータを用いた OV モデルに基づく最適速度とエネルギー削減,” 南山大学大学院数理情報研究科 2011 年度修士論文, 2012. (論文概要 <http://www.seto.nanzan-u.ac.jp/msie/ma-thesis/2011/KAWANO/m10mm040.pdf>) (accessed 2014-02-28)
- [11] 東京大学空間情報科学研究センター, CSIS-i 第 7 回シンポジウム「モバイルデータの社会的な有用性と期待」, 東京大学駒場第 II キャンパスコンベンションホール, 2014 年 1 月 28 日, 2014.

- [12] J. D. Ullman: データベース・システムの原理, 日本コンピュータ協会, 1985.
- [13] J. Vidya, C. W. Clifton, Y. M. Zhu: Privacy Preservation Data Mining, Springer, 2006.
- [14] Z. Yang, B. Yu, J. Lu, B. Yu, “An Adaptive Bus Arrival Time Prediction Model,” Proceedings of the Eastern Asia Society for Transportation Studies, Vol.7, pp.214-224, 2009.