

## 科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 5 月 30 日現在

機関番号：12608  
 研究種目：挑戦的萌芽研究  
 研究期間：2011～2012  
 課題番号：23656318  
 研究課題名（和文） 長期間の行動ログデータを用いた交通空間における行動文脈の抽出  
 研究課題名（英文） Estimation of travel contexts from longitudinal behavioural data  
 研究代表者  
 朝倉 康夫 (ASAKURA YASUO)  
 東京工業大学・大学院理工学研究科・教授  
 研究者番号：80144319

研究成果の概要（和文）：情報通信技術の高度化に伴って、GPS 付携帯電話を用いた調査システムによるデータや交通系 IC カードの料金収受データなどを利用することで、都市空間でのヒトの移動の高精度な観測が可能になりつつある。しかし、数ヶ月に渡る交通行動ログデータから、個人の長期的な交通行動特性を抽出する方法論は確立されてはいない。本研究では、自然言語処理の領域で研究されている文脈解析の理論を適用し、行動ログデータの文脈解析手法を開発することで、長期的な需要の変動の要因を把握するための方法を構築した。

研究成果の概要（英文）：Recent advances of information technologies enable us to obtain huge and continuous log data on travel behaviour. For example, a survey using GPS equipped mobile phone is often performed. Also the smart card systems used by the payment of transport systems collect huge data on travellers' usage of the transport systems. However, the methodologies of estimating travel contexts from such longitudinal log data are not still established. This study developed the methods to estimate travel contexts from the longitudinal behavioural log data, based on the methods of the natural language processing.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
交付決定額	3,000,000	900,000	3,900,000

研究分野：工学

科研費の分科・細目：土木工学、土木計画学・交通工学

キーワード：交通工学・交通行動分析・行動ログデータ

## 1. 研究開始当初の背景

情報通信技術の高度化に伴って、都市空間を移動するヒトの移動軌跡を高精度に観測できる方法の研究がわが国をはじめとする先進諸国で進められている。たとえば、プローブパーソン調査（以下 PP 調査）システムは、GPS 付携帯電話と WEB システムを組み合わせることで、ヒトの交通行動をシームレスかつ詳細に観測できるという特長を持っている。しかし、PP 調査で得られる、数週間から数ヶ月にわたる交通行動ログデータから、個人の長期的な交通行動特性を抽

出する方法論は確立されてはいない。

文脈解析(Context Analysis)は、主に自然言語処理の領域で研究されてきており、文字列の構文解析とその前後関係から意味内容を推定する意味解析から構成される。本研究では、言語と同様に、ヒトの交通行動もある種の文法と文脈から構成されていると仮定する。長期にわたる行動ログは自然言語に見られる一定の法則性とランダム性を内包しているとすれば、自然言語処理における文脈解析の考え方を適用した行動分析が可能になると考えられる。

## 2. 研究の目的

PP 調査等によって得られる長期間（数週間から数ヶ月）にわたる個人の交通行動ログデータの構造の特質を生かし、その行動文脈を解析するための方法論を開発することを目的とする。ここでいう行動文脈とは、個人の交通行動に内在する一定の行動ルールとその前後関係のことであり、長期間の行動ログに繰り返し現れる特徴的な行動パターンとその時空間分布を指す。本研究では、自然言語処理の領域で研究されている文脈解析の理論を適用し、行動ログデータの文脈解析手法を開発するものである。

## 3. 研究の方法

交通行動ログデータの文脈解析手法開発を目指し、平成 23 年度から 2 年間で以下の研究を行った（図 1）。

- (1) 長期間の交通行動分析、および自然言語処理分野の文脈解析手法に関する既往研究の整理
- (2) 長期間の行動ログデータからの分析用データベースの作成
- (3) ログデータの構造化モデル、ならびに行動文脈解析モデルの構築
- (4) モデルの改良と長期間行動ログデータへのモデルの適用
- (5) 研究成果のとりまとめ、および国内外の学会における研究成果の発表

(1)は、これまで交通工学分野の交通行動分析及び、自然言語処理の分野での文脈解析手法に関する研究内容を整理する。(2)では、既往の調査データから分析用データの抽出し、行動ログデータのデータベース化を行う。(3)では、非構造化データである行動データに、行動に関するラベリングを行い、行動文脈を把握するためのモデルの構築を行う。また、自然言語処理分野の方法を応用した行動ログデータの構文解析モデルの構築を行う。

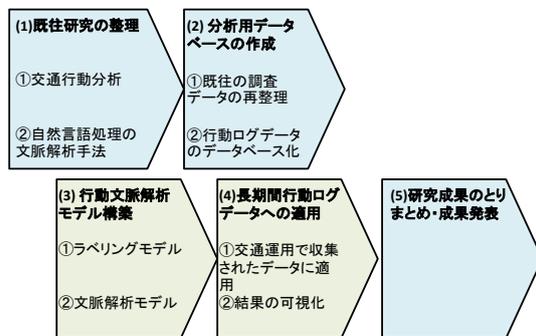


図 1 研究の方法

(4)では、実際の交通運用等で収集された、交通系 IC データや ETC データに(3)でのモデルを適用することで、構築したモデルが、長期観測データから行動変動の特徴抽出や、変動の解釈に役立つことを示すことができた。

## 4. 研究成果

本研究では、3. で述べたとおり、長期間の交通行動分析および自然言語処理分野の文脈解析手法に関する既往研究の整理を実施した上で、実際の交通系データに基づくデータベース構築、交通分野・自然言語解析分野のモデルを応用したモデル構築を行い、実データへの適用を行った。本研究の成果として、

- ログデータのラベリングモデル
  - ログデータの文脈解析モデル
- を構築することができた。これらのモデルを実際のデータに適用することにより、長期間、大量に収集された、交通系 IC カードデータや ETC データに記録された行動を解析し、需要の変動等の特徴をより高度に把握できることを確認できた。

### (1) ログデータのラベリングモデルの構築

#### ①概要

ログデータのラベリングモデルの構築を行った上で、実際の鉄道旅客のデータを収集した交通系 IC データに適用した。また、解析結果の可視化手法の開発を行った上で、研究成果のとりまとめと発表を実施した。〔雑誌論文〕②〔学会発表〕③

#### ②構築した手法

ログデータのラベリングモデルの構築では、単純ベイズ分類器に基づいたモデルを構築した。このモデルでは、過去の調査で収集された行動データであるパーソントリップ調査データ（以下 PT データと呼ぶ）を用いてベイズ確率モデルを構築した。この確率モデルを交通系 IC データに適用することで、交通系 IC データでは収集できない交通目的等の項目についてラベリングを行う手法となっている。

#### ③実データへの適用と可視化

IC データと PT データで共通するデータ項目は、乗降駅と乗降時刻である。そこで、本研究では、対象駅での降車時刻と、その駅で降車した時刻から次の乗車の時刻までの間隔（乗降間隔）を入力値として交通目的を推定する方法を構築した。

モデルの長期間行動ログデータへの適用では、都市部に路線をもつ鉄道会社 A 社 a 駅で観測された IC データを用いた。適用の対象期間は、2007 年 10 月 1 日から 2009 年 5 月

31日までの20ヶ月411日間の平日とした。この間に観測されたa駅を乗降するトリップ数は、6,913,979トリップであった。ベイズ確率モデルの推定に用いたPTデータは、2002年に実施された第4回京阪神パーソントリップ調査によるデータである。

図2は、構築したモデルを交通系ICデータに適用した結果の一例を示している。この図では、交通目的毎のa駅で降車を行ったトリップを日毎に集計している。図より、トリップ目的毎に日変動の特徴が異なることが把握できる。例えば、この駅では、通勤目的のトリップが多く、帰宅、私用、業務、通学の順に続くことがわかる。私用以外の目的のトリップでは、お盆や年末年始にトリップ数が減少していることがわかる。帰宅目的のトリップは他の目的のトリップに比べて、4月から9月にかけてのばらつきが大きくなっている。これは、この路線の他の駅近辺にあるイベント施設で行われているイベントの期間に一致しており、その帰宅の乗客が他の社局の路線への乗り換えているものを捉えたものだと推察できる。

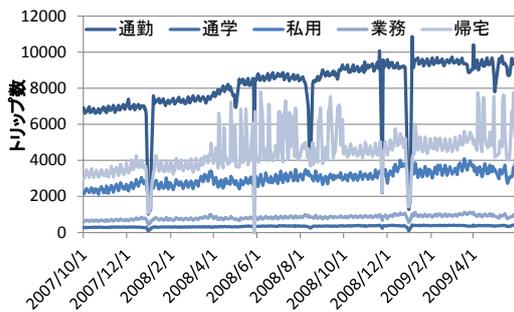


図2 トリップ目的の推定結果

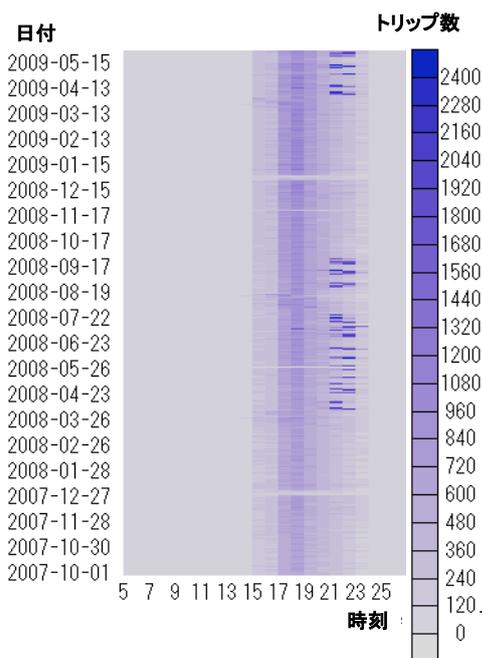


図3 時間帯別の推定結果 (帰宅目的)

図3は、解析結果に可視化技術を適用したものの一例である。可視化技術を適用することより、各トリップの日ごとの時系列変動をより詳細に把握できることがわかる。例えば、帰宅目的のトリップは、4月から9月にかけて21時以降にトリップ数の大きな増加が見られる日がある。この増加は、図2の帰宅目的のトリップで見られた大きなばらつきが発生する日と一致しており、イベントからの帰宅交通が中心であることと推察される。

(2) ログデータの文脈解析モデル

①概要

行動ログデータの文脈解析モデルでは、従来の交通行動分析で用いられている行動モデルを改良することでログデータに含まれる行動文脈の解釈を行う方法を構築するとともに、自然言語解析に用いられる隠れマルコフモデルを応用することで、鉄道利用者の行動文脈を分析するための統計モデルを構築した。このモデルを交通系ICデータに適用することで、交通行動の解釈をデータに収集された行動パターンから行えることが示された。また、研究成果のとりまとめと国際ワークショップでの発表を実施した。〔雑誌論文〕

①③ [学会発表] ①②④⑤)

②構築した手法

ログデータの文脈解析モデルの構築では、隠れマルコフモデルに基づいた特定のODを利用する利用者に関するDay-to-dayの鉄道利用についてその行動文脈を解釈するモデルを構築した。このモデルでは、鉄道利用者の行動の変動には、「短期的」な変動と「長期的」な変動の2種類があると仮定し、短期的な変動として曜日ごとの利用確率を定義した。短期的な行動のパターンが変わることによって長期変動が起こるとした上で、この変動が隠れマルコフモデルに従っていると仮定し、モデルを記述した。

③実データへの適用

鉄道会社AのODペアN駅・S駅で2007年10月22日～2009年5月31日の588日間に観測された利用データを用いて分析を行った。期間中の利用者数は、78,246人であり、延べ646,559人日の利用があった。

分析では、EMアルゴリズムの一種であるBaum-Welchのアルゴリズムによってモデルパラメータの推定を行った後、Viterbiのアルゴリズムを全利用者のデータに適用し、行動文脈を推定した。

図4は、行動の短期的な変動である曜日ごとの利用確率をモデルによって抽出された異なるグループ毎に示したものである。グループ1は、ほとんど利用の無いグループであり、グループ4まで徐々に利用確率が高いグ

グループとなっている。グループ3、グループ4では、平日の利用確率が高く、これに比べると、土曜、休日の利用確率が低くなっていることから、行動文脈が異なると推察できる。

図5は、各グループの所属人数を求め、グループ2～4の所属人数と各日の利用人数を図示したものである。今回の分析の全利用者数は、78,246人であることから、約9割の利用者は、ほとんど利用がないグループであるグループ1に属していることがわかる。さらに、中頻度、高頻度のグループであるグループ3、4に所属する利用者数は、それぞれ1%程度であることがわかる。2009年の3月と4月を見ると、低頻度の利用者が短期間に集中して利用する時期があることがわかる。これは、沿線で開催された何らかのイベントと関連しているものと推察できる。また、グループ3、4のような中頻度、高頻度の利用者層については、比較的構成人数が安定していることがわかる。



図4 曜日利用パターンの推定結果

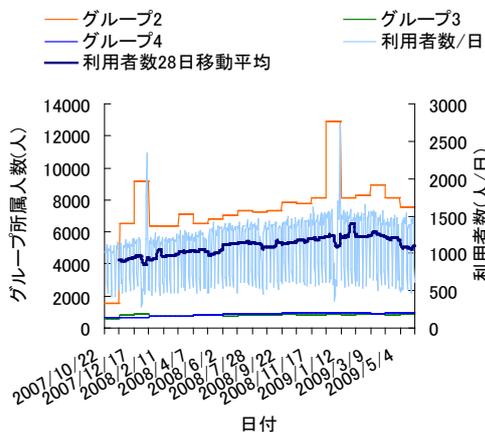


図5 グループ所属人数の推定結果

### (3) 成果のまとめ

ログデータのラベリングモデルの構築では、データに含まれる交通需要変動をトリップ目的ごとの変動として解釈するために、ICデータとPTデータをベイズ推定により融合し、分析する手法を構築した。ログデータの文脈解析モデルでは、従来の行動モデルに加え自然言語の解析に用いられる隠れマルコ

フモデルを応用することで、利用者の変動を分析する手法を構築した。また、これらの方法に可視化技術を適用することにより、変動の要因をより直観的に捉えることが可能となった。これらの成果により、個人の交通行動に内在する一定の行動ルールとその前後関係である行動文脈を長期間の行動ログから特徴的な行動パターンとして抽出するための一手法を構築することができた。

今後、このようなシステムを活用することで、従来の可視化手法と比較して、より詳細に鉄道沿線のイベントや集客施設による影響などによる変動の要因を推測することができると考えられる。また、特異な変動が見られた場合には、変動があった利用者の利用目的を考慮した上で、より詳細な調査を検討するとともに、改善施策やマーケティングの検討への展開も見込めるだろう。

### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計3件)

- ① 柳沼秀樹, 石田貴志, 足立智之, 朝倉康夫; 大規模データを用いた情報提供と経路選択行動の実証的分析. 第32回交通工学研究発表会論文集, 査読有, pp. 375-379, 2012
- ② Yasuo ASAKURA, Takamasa IRYO, Yoshiki NAKAJIMA, Takahiko KUSAKABE; Estimation of Behavioural Change of Railway Passengers Using Smart Card Data. *Public Transport*. 査読有, 4(1), pp. 1-16, 2012
- ③ 上田大樹, 井料隆雅, 朝倉康夫; 都市高速道路のランプ間OD交通量に旅行時間が与える影響の実証分析. 第31回交通工学研究発表会論文集, 査読有, pp. 147-152, 2011

[学会発表] (計5件)

- ① Takamasa IRYO, Hiroki UEDA, Kohei OZASA and Yasuo ASAKURA; Empirical Analysis of Demand Fluctuation Caused by Incidental Congestion. *Proceedings of the 5th International Symposium on Transportation Network Reliability*, 査読有, 2012年12月19日, pp. 991-1010, Hong Kong, China
- ② 上田大樹, 井料隆雅, 朝倉康夫; ETC統計データを用いた旅行時間がランプ間OD交通量に与える影響の実証分析, 第46回土木計画学研究発表会. 2012年11月3日, さいたま(埼玉大学)

- ③ 日下部貴彦, 朝倉康夫; データフュージョンによる行動データマイニングのための基礎分析. 第45回土木計画学研究発表会, 2012年6月1日, 京都(京都大学)
- ④ Takahiko KUSAKABE, Yasuo ASAKURA, Behavioural Data Mining for Railway Travellers with Smart Card Data. *2nd International Workshop on Traffic Data Collection & its Standardisation*, 2011年9月23日, Brisbane (Queensland University of Technology), Australia
- ⑤ 日下部貴彦, 辻本洋平, 朝倉康夫; 旅行時間信頼性情報による高速道路利用者の行動変化の分析. 第43回土木計画学研究発表会, 2011年5月29日, つくば(筑波大学)

[その他]

ホームページ

<http://www.plan.cv.titech.ac.jp/asakura/lab/>

## 6. 研究組織

### (1) 研究代表者

朝倉 康夫 (ASAKURA YASUO)  
東京工業大学・大学院理工学研究科・教授  
研究者番号: 80144319

### (2) 研究分担者

井料 隆雅 (IRYO TAKAMASA)  
神戸大学・工学(系)研究科(研究院)・教授  
研究者番号: 10362758

日下部 貴彦 (KUSAKABE TAKAHIKO)  
東京工業大学・大学院理工学研究科・助教  
研究者番号: 80604610