

論文 / 著書情報  
Article / Book Information

論題(和文)	対話して学習するプローブパーソン調査システムの提案
Title	
著者(和文)	後藤 啓人, 定金 乾一郎, 瀬尾 亨, 日下部 貴彦, 朝倉 康夫
Authors	Hiroto Gotoh, Kennichiro Sadakane, Toru Seo, Takahiko Kusakabe, Yasuo Asakura
出典 / Citation	第13回ITSシンポジウム2015, , ,
Citation(English)	, , ,
発行日 / Pub. date	2015, 12
Note	本著作物は第13回ITSシンポジウム2015の発表論文である

# 対話して学習するプローブパーソン調査システムの提案

後藤啓人<sup>\*1</sup> 定金乾一郎<sup>\*1</sup> 瀬尾亨<sup>\*1</sup> 日下部貴彦<sup>\*1</sup> 朝倉康夫<sup>\*1</sup>  
東京工業大学<sup>\*1</sup>

本研究は、スマートフォンなどの移動体通信機器を用いた、対話型交通行動調査手法の枠組みを提案する。提案手法は、長期間の調査での回答者の回答頻度の低減を意図したものである。本手法では、オンラインでの交通目的推定モデルと学習モデルを対話プロセスとして構成している。推定モデルは、交通目的の推定を調査期間中に自動的に行うためのものである。学習モデルは、回答者の回答に基づいて推定モデルを更新するものである。過去のプローブパーソン調査で収集したデータを使用したシミュレーション分析により、提案手法の精度を検証した。その結果、一定の誤判定を許容する場合に提案手法は交通行動調査における質問数減少に寄与し得ることが示された。

## Development of Interactive and Self-Learning Probe Person Survey System

Hiroto Goto<sup>\*1</sup> Kenichiro Sadakane<sup>\*1</sup> Toru Seo<sup>\*1</sup> Takahiko Kusakabe<sup>\*1</sup> Yasuo Asakura<sup>\*1</sup>  
Tokyo Institute of Technology<sup>\*1</sup>

This study proposes a framework of an interactive activity-travel survey method implemented on mobile devices such as smartphones. The proposed method was developed intended to reduce frequency of questions for a respondent in long-term behavioural survey. The method employs an online trip purpose estimation model and an online machine learning method as interactive processes. The estimation model is for automatically estimating a trip purpose during a survey and the online machine learning method is used for updating the estimation model based on answers from a respondent. The proposed method was examined by a simulation using the data obtained by a previous probe person survey. The results showed that the method can reduce the number of questions in the behavioural survey when the analysts can accept a certain amount of failure estimation.

**Keyword: Travel-Activity Survey, Online Machine Learning, Online Travel Context Estimation**

### 1. はじめに

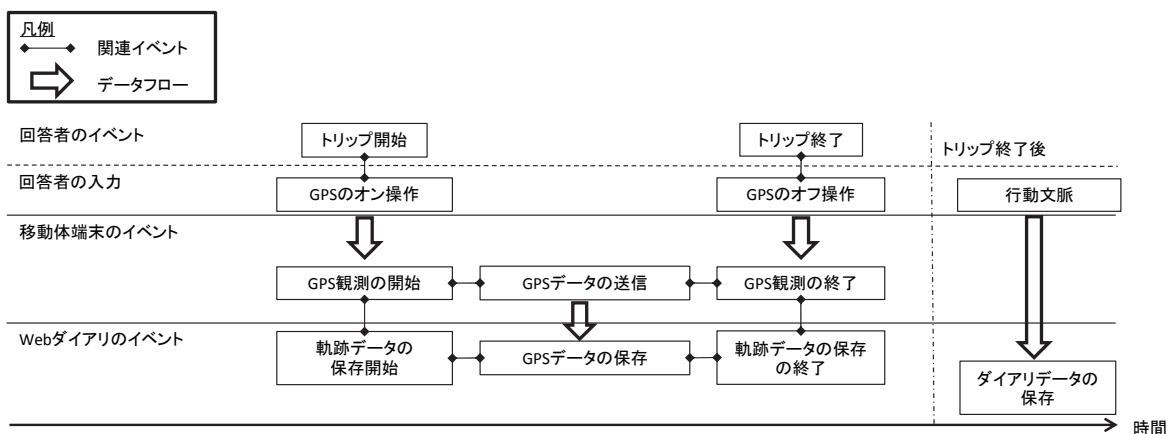
1990年代後半より、交通行動データを効率的に収集するための方法として移動体通信機器を用いたPP（プローブパーソン）調査などの調査手法が開発

されてきた<sup>1)</sup>。このような調査手法では、GPS (global positioning system) 等で収集された回答者の移動軌跡が通信機器によって自動的に収集される。従来のPP調査では、移動軌跡データを自動的に収集するとと

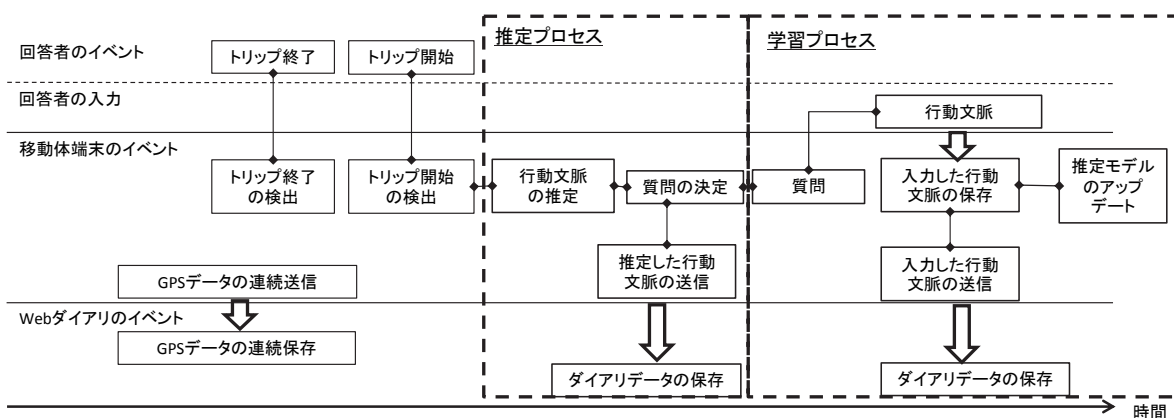
もに、Web 上での交通ダイアリ調査 (Web ダイアリ調査) を組み合わせ、軌跡データから直接的に判別が難しい交通行動の文脈に関わる調査項目 (例: 交通目的、交通手段) についての情報も収集する。これにより、PP 調査では、従来の PT 調査 (パーソントリップ調査) に代表される紙媒体を用いて行う調査と比較して、長期間かつ時空間的に高解像度な調査が可能になった。しかし、PP 調査の枠組みでも、回答者が、Web ダイアリ調査を用いて、交通行動の文脈に関わる詳細な情報をすべて入力しないといけないという点では、従来型の PT 調査と変わりはない。したがって、このような負担により月単位・年単位と言った長期間の調査が難しいという点では、従来の PT 調査と変わりがないといえる。既往の文献でも、これらを原因とする長期間の行動調査の問題点が示されており<sup>2)3)</sup>、移動体通信機器を用いる PP 調査であったとしても、既往の交通行動調査での回答者数は、千人以下であり、調査期間は 2, 3 か月以下にとどまっている<sup>1)4)5)</sup>。

近年では、通信機器等から得られるデータを利用し、交通目的や交通手段などの行動文脈を自動的に補完するための手法が開発されてきている<sup>6)7)8)9)</sup>。これら手法では、行動文脈を補完するための判別関数のパラメータを調査に先立って推定しておく必要がある。すなわち、行動文脈と移動軌跡等の観測値を結びつけるためのデータを行動調査に先立って収集する必要があることを意味しており、独立した調査が困難である。さらに、行動文脈と実際の行動結果である移動軌跡の関係は、回答者のライフスタイルや周辺環境によって左右される可能性がある。したがって、より長期間の調査では、調査期間中の回答者の状況によってこの関係が決定され、更新されるのが望ましいと考えられる。

本研究は、スマートフォン等の移動体通信機器を用いた対話型交通行動調査手法の枠組みを提案する。事前情報に依存せずに長期間の交通行動調査に対応するために、本手法では、オンラインでの交通行動推定モデルと対話プロセスによる学習モデルを採用



(a)従来のプローブパーソン調査システム



(b)オンライン機械学習法によるプローブパーソン調査システム

図1 トリップ目的から設問までのアクティビティ図

する。推定モデルは、交通目的の推定を調査期間中に自動的に行うために用いる。学習モデルは回答者の回答に基づいて推定モデルを更新するのに用いる。提案手法では、GPS等で自動的に収集されるデータを用いてシステムが自動的に交通目的等の行動文脈を推定し、推定結果の信頼度に応じて、実際の行動文脈に関する質問を生成する頻度を調節する。実際の行動文脈を質問した場合には、回答者の返答に応じて、推定モデルを更新する。このことにより、従来の交通行動調査手法と比較して提案する手法では、回答者が要求される情報の入力頻度を低減できることが期待される。また、従来のオフラインでの行動文脈の推定方法と比較して、提案する手法は事前情報への依存が小さい。加えて、回答者の現在の状況に応じて推定モデルを更新できることから、長期間調査中でのライフスタイルや周辺環境の変化にも対応できると期待できる。

## 2. 方法論・モデル

図1は、従来型のPP調査と提案する調査システムでの、回答者とシステムの動作プロセスを示している。図1(a)のように、従来のPP調査手法では、回答者は、トリップの開始・終了時に機器を操作する必要があるほか、トリップ終了後にWebでのダイアリーを用いてすべての行動文脈を記入する必要がある。本研究では、行動文脈は移動軌跡データから直接的に得られない特性、トリップ属性は移動軌跡データの処理のみで得られる特性と定義する。例えば、行動文脈は、交通手段や交通目的であり、トリップ属性はトリップの出発地、到着地、出発時間、到着時間である。なお、本研究では、行動文脈として交通目的を推定するものとして定式化を行う。

図1(b)に示すように、本研究での提案手法は、推定プロセスと学習プロセスの二つの過程で構成される。この手法は、移動体端末から得られる移動軌跡のオンラインデータから、移動滞在識別法<sup>7)</sup>などを用いて、トリップを検出した上で適用することを想定している。つまり、トリップの出発・到着地、出発・到着時間に関しては、自動的に識別されていることを前提としている。推定プロセスでは、これらの識別されたトリップの交通目的を推定する。これにより、従来の調査と比較し、回答者のダイアリー入力頻度が低減されると期待される。学習プロセスでは、回答者が実際の交通目的について回答を行った場合に、回答結果に基づいて推定モデルを回答者ごとにキャリブレーションする。

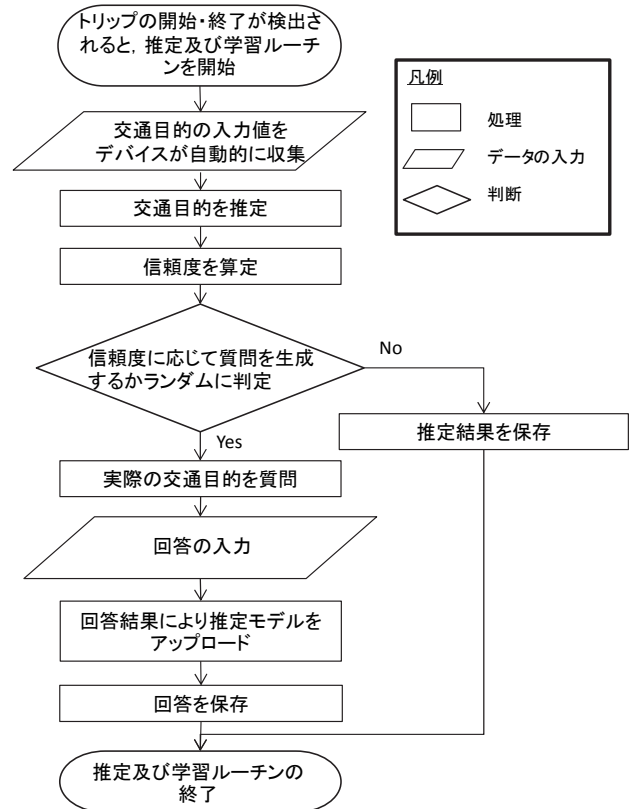


図2 推定と学習のフローチャート

図2に推定と学習のフローチャートを示す。交通目的は、回答者のトリップの終了を検知後に推定される。交通目的に関する回答者への質問は、推定結果の信頼度に応じた頻度でランダムに生成される。つまり、信頼度が十分に高い場合には、回答者は交通目的に関する回答を頻繁には求められなくなる。質問への回答が得られた場合には、回答結果に基づき推定モデルを更新する。

本研究では、推定モデルと学習モデルに、最も単純なモデルの一つである単純ベイズ分類器<sup>10)</sup>を採用する。単純ベイズ分類器は、他の機械学習モデルと比較してモデルのキャリブレーションに必要なデータ数が少ないという特長がある。提案する手法は、回答者ごとに推定モデルが必要であるため、推定モデルのキャリブレーションに用いることができるデータ数は比較的少ないと想定される。例えば、1日に平均3トリップする回答者のデータは、月に90トリップ分しか収集できない。

本研究のモデルでは、交通目的の推定対象であるトリップの次のトリップの出発時に利用できる情報を使用する。すなわち、推定対象であるトリップの後にいった活動内容である前のトリップの交通目的

を推定するものである。回答者は、次のトリップを行っている間に、推定対象のトリップについての質問に回答する。2-1 節に単純ベイズ分類器による交通目的の推定手順を示す。2-2 節には質問を提示する条件と学習手順を示す。

## 2-1 推定プロセス

推定モデルは、学習プロセスによって推定される確率関数によって構成される。cを交通目的の集合Cの要素とし、集合Cは、

$$C = \{\text{通勤, 帰宅, 業務, 私用, その他}\} \quad (1)$$

とする。

Kをトリップ属性の集合とし、

$$K = \left\{ \begin{array}{l} \text{平日, 到着時刻, 出発地, 目的地,} \\ \text{活動時間} \end{array} \right\} \quad (2)$$

とする。式(2)のトリップ属性は、それらの値が交通目的によって特徴付けられるものであると期待されることから選択した。「平日」は、平日とそれ以外での行動が異なることを想定して設定した。「出発地」・「目的地」は自宅と職場、その他を示す変数であり、自宅の位置については、あらかじめ調査登録の過程で登録されていると想定している。到着時刻と到着地は、トリップ後の活動を開始した時間とその場所とみなすことができる。活動時間は、交通目的の推定対象のトリップの到着と次のトリップの出発までの時刻の差で定義する。これらトリップ属性について、 $x_k$ をトリップ属性 $k \in K$ の値とする。また、 $x_k$ が取り得る値は、集合 $X_k$ に含まれる離散的な値である。集合 $X_k$ の定義は以下の通り。平日に関する属性は、 $X_{\text{平日}} = \{\text{平日, その他}\}$ とする。到着時刻に関する属性に関して、 $X_{\text{到着時刻}}$ は3時間ごとの時間帯で離散化した。出発地、目的地それぞれに関する属性は $X_{\text{出発地}} = X_{\text{到着地}} = \{\text{自宅, 職場, その他}\}$ とする。活動時間に関する属性に関して、 $X_{\text{活動時間}}$ は1時間ごとの時間帯で離散化した。また、12時間を超えたトリップは、12時間以上として分類した。

単純ベイズ分類機では確率関数 $P(c)$ と条件つき確率 $P(x_k|c)$ を用いる。本モデルでは、 $P(x_k|c)$ が $k$ 毎にそれぞれ独立である仮定する。トリップ属性の値 $x = \{x_k|k \in K\}$ が観測されたとき、そのトリップの交通目的が $c$ である確率は、

$$P(c|x) = \frac{1}{P(x)} \prod_{k \in K} P(x_k|c)P(c) \quad (3)$$

と表すことができる。 $P(x)$ は、トリップ属性の生起

確率である。なお、 $x$ が与件である場合に、式(3)を最大にする交通目的 $\hat{c}$ を推定したい場合には、 $P(x)$ は一定値とみなせるため、推定結果に関与しない。

よって、推定する交通目的は

$$\hat{c} = \arg \max_{c \in C} \prod_{k \in K} P(x_k|c)P(c) \quad (4)$$

と表すことができる。

## 2-2 学習プロセス

回答者への質問は、交通目的の推定結果の信頼度に応じて、確率的に発生させる。具体的には、推定結果の誤判定率が、設計過程であらかじめ設定した許容誤判定率 $P_{af}$ になるように、質問を確率的に発生させる。なお、許容誤判定率は、調査を通しての誤判定率と等しくなると期待される。

信頼度を、式(4)で推定した交通目的 $\hat{c}$ を用いて

$$P(\hat{c}|x) = \frac{\prod_{k \in K} P(x_k|\hat{c})P(\hat{c})}{\sum_{c \in C} \prod_{k \in K} P(x_k|c)P(c)} \quad (5)$$

と表すとき、質問の発生確率は、信頼度と許容誤判定率 $P_{af}$ を用いて、

$$P_q = \max \left\{ 0, 1 - \frac{P_{af}}{1 - P(\hat{c}|x)} \right\} \quad (6)$$

と定義する。

上記のプロセスで質問が発生しない場合や回答者による回答がない場合には、推定結果は式(4)の $\hat{c}$ と確定し、学習プロセスで推定モデルの確率関数の更新を行わない。質問が発生し、回答者が回答した場合には、回答結果を教師データとして確率関数を更新する。 $r = (t_a, c_a, x_a)$ を回答結果を表す集合とする。この時のトリップ属性は $x_a = \{x_{ka}|k \in K\}$ 、 $c_a$ は交通目的の回答結果、 $t_a$ は回答した時刻である。この回答結果を用いて、時刻 $t$ に $P(x_k|c)$ を求める式は、

$$P(x_k|c) = P_{c x_k} = \frac{\sum_{r \in A_t} \delta(r, c, x_k)}{\sum_{x_k \in X_k} \sum_{r \in A_t} \delta(r, c, x_k)} \quad (7)$$

と表すことができる。なお、

$$\delta(r, c, x_k) = \begin{cases} 1 & \text{if } c_a = c \cap x_a = x \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

であり $A_t$ は、時刻 $t$ より前の回答履歴の集合である。

式(7)は、交通目的の回答が得られた場合に、トリップ属性 $k$ の値が $x_k$ であった割合である。同様に、 $P(c)$ は、

$$P(c) = P_c = \frac{\delta(r, c)}{\sum_{r \in A_t} \delta(r, c)} \quad (9)$$

と表すことができる。ただし、

$$\delta(r, c) = \begin{cases} 1 & \text{if } c_a = c \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

である。

### 3. 実証分析

#### 3-1 検証法

過去の PP 調査で得た行動データを用いたシミュレーション分析により、提案手法を検証する。なお、分析に用いるデータを収集した調査は、従来型の PP 調査であり、本研究で提案した対話・学習を組み込んだ調査ではない。使用するデータは、2007年に松山市にて実施された PP 調査で得たものであり、Web ダイアリ上で回答者が記入した交通目的、出発・到着地、トリップ開始・終了時刻で構成されている。回答者は、2007/10/29 から開始した 58 人、2007/10/17 から開始した 138 人の合計 196 人である。調査期間は最大 14 日間であり、開始日から数えて 6 日、7 日、13 日、14 日が週末である。また、期間中の月～金に祝日はなかった。回答者は、期間中に延べ 8104 トリップを行った。なお、これは 41.3 トリップ/人に相当し、3.83 トリップ/人日に相当する。

分析に用いるデータには、Web ダイアリで得られた全てのトリップに対する交通目的が含まれていることから、この交通目的を用いて学習プロセスを再現し、推定モデルを求めた。すなわち、2-2 節で示した質問が発生する場合には、データに含まれる交通目的を質問に対する回答とみなし、学習プロセスに使用した。質問が発生しない場合には、推定結果の正誤にかかわらず、推定結果を交通目的として使用した。学習プロセスで質問発生判断の際に使用する許容誤判定率  $P_{af}$  は、5% とした。なお、これは、シミュレーション分析では、5% の誤判定が想定されることを示している。

#### 3-2 結果

図 3 は、シミュレーション分析の結果である。横軸は、各回答者の調査開始日からの日数である。縦軸は質問率、成功率、誤判定率である。ここで、質問率とは、回答者に交通目的を問う質問が発生したトリップの割合である。成功率とは、質問が発生せず、かつ交通目的を正しく推定できたトリップの割合である。誤判定率は、質問が発生せず、かつ交通目的を誤って推定したトリップの割合である。

全体的な傾向として、調査の日数経過とともに質問率が減少し、成功率が上昇している。平日の場合、最終的な質問率は約 40% となった。休日である第 6, 7, 13, 14 日は平日と比較し質問率が高い。これは、休日のサンプルが少ないこと、および休日のトリップが不規則であることが原因と考えられる。一方、誤判定率は全期間通して 7% 程度と一定であり、交通目

的推定の精度が保たれていることがわかる。ただし、その平均値は、事前に与えた  $P_{af} = 5\%$  よりもわずかに大きい。これは単純ベイズモデルでの独立性の仮定の影響が原因と考えられる。以上により、本手法は、回答者の情報に基づき適応的に推定モデルを学習することで、回答負担を低減できたといえる。

図 4 に、交通目的ごとの推定結果を示す。結果は、調査の 1 週目と 2 週目に分けて集計した。すべての交通目的の 2 週目の質問率は 1 週目に比べ減少している。質問率の最小値は通勤であった。この質問率は 40.3% から 24.9% に減少した。これは通勤が規則的に行われていて、そのトリップのほとんどが自宅と職場間であったからだと考えられる。帰宅での質問率は、1 週目が 52.6%、2 週目が 34.7% であった。これは、帰宅には通勤に比べて様々な出発地があるからだと考えられる。通勤と帰宅とは対照的に業務、娯楽、その他での質問率に大きな改善は見られなかった。この理由は、これらのトリップには決まった出発地、到着地、時間、期間がないためと考えられる。誤判定率にも同様の傾向がある。娯楽とその他の誤判定率は、ほかの交通目的よりも高いことが示されている。この結果は、これらの交通目的での活動には、トリップ属性に強い規則性がないことを示唆している。

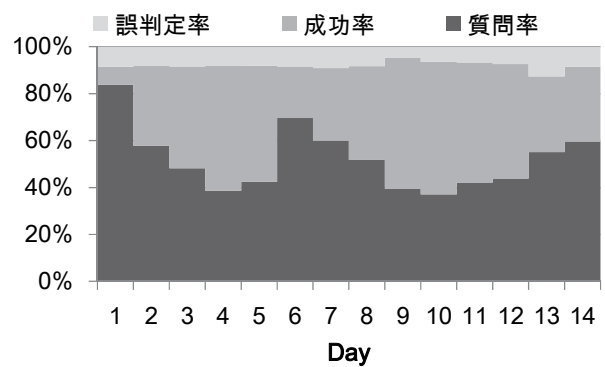


図 3 ウェブダイアリデータを用いた検証結果

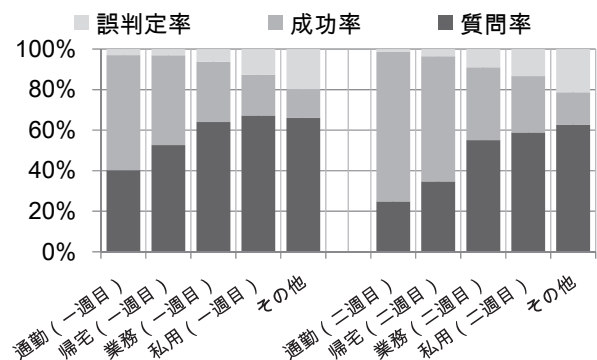


図 4 交通目的別の推定結果



#### 4. 結論

本研究では、行動文脈の一つである交通目的を推定するモデルを、調査期間中の対話プロセスで更新可能な、対話して学習する交通行動調査手法を提案した。提案した手法は、推定結果の信頼度が十分でないときに、回答者に行動目的を問う質問を発し、その結果に基づきモデルを更新する設計となっている。これにより、本手法は、事前情報に依存せず、被験者の回答負担が小さく、かつ一定程度の精度が担保された行動調査が可能と期待される。

従来 PP 調査で得た実際の行動データを使用したシミュレーション分析により、提案手法についての精度や妥当性を検証した。この結果、取得データの精度を維持したまま、回答者による記入の頻度を減少できることが示された。

今後の解析では、トリップの自動検出（移動滞在識別手法）と提案手法を実装したスマートフォンを使用した実地調査によって、提案手法を検証する。この調査により、提案手法が、実際の長期間の行動調査に必要な性能を満たしているかについて明らかにする予定である。

#### 5. 注釈

本稿は、著者らが The 14th International Conference on Travel Behaviour Research (IATBR2015)で発表した「Interactive Online Machine Learning Approach for Activity-Travel Survey」に基づいている。

#### 参考文献

- 1) 朝倉康夫, 羽藤英二, 大藤武彦, 田名部 淳, 2000. PHS による位置情報を用いた交通行動調査手法, 土木学会論文集 (653), 95-104
- 2) Kitamura, R., 1990. Panel analysis in transportation planning: An overview. *Transportation Research Part A: General* 24, 401–415.
- 3) 内田敬, 飯田恭敬, 1993. 交通行動パネル調査の方法論的検討, 土木計画学研究・論文集 (11), 319-326.
- 4) Wolf, J., Guensler, R., Bachman, W., 2001. Elimination of the travel diary: an experiment to derive trip purpose from GPS travel data. *Transportation Research Record* 1768, 124–134.
- 5) Draijer, G., Kalfs, N., Perdok, J., 2000. Global positioning system as a data collection method for travel research. *Transportation Research Record* 1719, 147–153.
- 6) Shen, L. and Stopher, P. R., 2013. A process for trip purpose imputation from Global Positioning System data. *Transportation Research Part C: Emerging Technologies* 36, 261–267.
- 7) 塚本健太郎, 佐藤仁美, 森川高行, 2014. GPS やコンテキストデータを活用した行動目的の推定手法に関する研究, 土木計画学研究・講演集 50,
- 8) Kusakabe, T. and Asakura, Y., 2014. Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies* 46, 179–191.
- 9) 中村英夫, 宮下浩一, 羽藤英二, 岸井隆幸, 2015. パーソントリップ調査のための加速度センサを活用した Random Forest による移手段推定手法, 交通工学論文集 1(5), 10-18.
- 10) Rish, I., 2001. An empirical study of the naïve Bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Seattle, Washington, USA, 4–10 August 2001.