Interactive Online Machine Learning Approach for Activity-Travel Survey

Toru Seo[a,*], Takahiko Kusakabe[b], Hiroto Gotoh[c], Yasuo Asakura[a]

[a]*Tokyo Institute of Technology, 2-12-1-M1-20, O-okayama, Meguro, Tokyo 152-8552, Japan*
[b]*Center for Spatial Information Science, at the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan*
[c]*Ministry of Land, Infrastructure, Transport and Tourism, Japan*

Abstract

This article proposes a framework for an interactive activity-travel survey method, implementable on mobile devices such as smartphones. The proposed method was developed to reduce the burden (i.e., frequency of questions) on respondents in long-term behavioral surveys, without relying on external data sources. The method employs an online travel context estimation model and an online machine learning method as interactive processes. The estimation model is used for automatically estimating travel contexts during surveys, while the online machine learning method is used for dynamically updating the estimation model, based on answers from respondents. The proposed method was examined by simulations using data obtained from a past probe person survey. The results suggest that the frequency of inputs by respondents in surveys can be significantly reduced, while maintaining high accuracy of the obtained data. For example, the method successfully estimated certain types of trips (e.g., commuting) and the behaviors of certain respondents (e.g., those whose activity-travel pattern is recurrent) because of the learning process and reduced survey burden on them. Meanwhile, although the method could not always precisely estimate some other types of trips, it eventually obtained accurate results because of the interaction process. Therefore, the proposed method could be useful to reduce the burden on respondents in long-term surveys, while maintaining high data quality and capturing traveler heterogeneity.

Keywords: activity-travel survey, behavioral context inference, traveler heterogeneity, GPS, smartphone, naïve Bayes classifier

1. Introduction

Activity-travel survey methods with tracking devices have been developed since the late 1990s as effective methods to collect behavioral data. In these surveys, trajectories of survey respondents are automatically collected by mobile instruments such as the Global Positioning System (GPS). Internet web-based diaries, synchronized with the data from mobile instruments are used to complement the detailed information on trips and activities. Compared to traditional surveys like person trip surveys and paper-based diary surveys, the mobile instruments improve the observation period and resolution in both space and time dimensions significantly (Murakami and Wagner, 1999).

However, even if such mobile instruments are applied to a survey, survey respondents are still required to manually input the detailed activity-travel information because the information obtained from the instruments does not directly contain activity-travel attributes and behavioral contexts. Typical examples of such behavioral contexts are the trip purpose (or equivalently, activity type), travel mode, transport-related fare, and travel companion(s). Therefore, much time and effort is still required for the respondents as the survey period becomes longer. As a consequence, the number of respondents in most tracking surveys remains less than a thousand, and survey duration, less than a few months (e.g., Murakami and Wagner, 1999; Draijer et al., 2000; Wolf et al., 2001; Asakura and Hato, 2004). These situations show the difficulties associated with collecting day-to-day data for continuous long-term periods via these

*Corresponding author
Email addresses: t.seo@plan.cv.titech.ac.jp (Toru Seo), t.kusakabe@csis.u-tokyo.ac.jp (Takahiko Kusakabe), asakura@plan.cv.titech.ac.jp (Yasuo Asakura)

surveys. Reasons for such difficulties include costs, processing load, inaccuracy, and privacy protection of respondents (Kitamura, 1990; Axhausen et al., 2007). In summary, it is still difficult to collect large-scale individual activity-travel data for a long duration, although such data is very useful to estimate long-term activity-travel behavior (Kitamura et al., 2003; Arentze and Timmermans, 2009; Bhat et al., 2013).

To resolve these problems, many existing studies have attempted to automatically complement behavioral contexts to the data obtained from mobile instruments and passive data collection methods. In this article, trip purpose inference is considered particularly, because of its importance and the associated current research gaps. Note that trip purpose inference is almost equivalent to activity type inference. In existing studies, Auld et al. (2009) proposed a prompted recall survey assisted by GPS and activity inference. Shen and Stopher (2013) developed a trip purpose imputation method for GPS data based on household travel survey data. Cottrill et al. (2013) and Kim et al. (2014) developed methods to automatically estimate travel attributes on a web-based diary system from mobile instrument and historical behavioral data. Hasan et al. (2013) and Hasan and Ukkusuri (2014) proposed activity pattern classification methods using geo-location data in online social media. Oliveira et al. (2014) developed a trip purpose identification method based on GPS and a discrete choice model to incorporate various types of information. Washington et al. (2014) applied a Bayesian inference method based on small-scale survey data to impute non-chosen attributes in large-scale data. Kusakabe and Asakura (2014) and Kusakabe et al. (2016) developed a trip purpose estimation method for transit smart card data based on the person trip survey data and naïve Bayes classifier, and applied in Japan and Australia. Alexander et al. (2015) presented methods to estimate number of trips and their purposes from call digital record and census data. Siripirote et al. (2015) proposed use of plate scanning and traffic count data to calibrate an activity-based model. The notable feature of this method is that it is not based on sampling nor offline information; however, it is only applicable to trips with personal cars. Xiao et al. (2016) developed a trip purpose estimation method using land-use data and GPS traces. Han and Sohn (2016) also developed an activity imputation method using land-use data and smart card data. For comprehensive reviews on these topics, see Shen and Stopher (2014), Rasouli and Timmermans (2014), Gong et al. (2014), and Prelipcean et al. (2017).

The aforementioned existing trip purpose inference methods (except for Siripirote et al. (2015)) depend on offline common information (e.g., land-use, census data) and/or predetermined parameters of discriminant functions (e.g., behavioral model) for behavioral contexts (e.g., trip purpose) which model relations between behavioral contexts and observable real-time information (e.g., GPS traces). Although the use of such predetermined relations is beneficial to reduce the survey burden on respondents, some limitations could arise as a result. First, preceding data acquisition is required to derive and calibrate the relations. Second, traveler heterogeneity may be ignored; in reality, the relations can vary among travelers and change dynamically, depending on lifestyles and environment. Third, the accuracy of inference during actual surveys is difficult to be guaranteed or assessed. These limitations can be problematic for certain surveys, such as those conducted after significant changes in the built environments, those intended to consider diversity in the society, and those conducted in developing countries.

The aim of this article is to develop a mobile-phone-based behavioral context inference method that requires no preceding data acquisition, and captures traveler heterogeneity and dynamical behavioral changes. To achieve this, this article proposes a framework for an interactive activity-travel survey method implementable on mobile devices such as smartphones. In order to adapt the method for long-term activity-travel surveys with less burden, the method employs an *online behavioral context estimation model* and an *online machine learning method* with an *interactive process*. The estimation model is for automatically estimating behavioral contexts (specifically, trip purpose) during surveys. The learning method is used for updating the estimation model to adapt the model to observed activity-travel behaviors based on interaction with the respondents. The interactive process between a survey system and a respondent is continued during the survey period, as explained by the following. The system automatically estimates a behavioral context, and occasionally raises questions for an actual behavioral context depending on the confidence level of the estimation results. Based on the answer of the respondent, the method updates the estimation model. Comparing with conventional travel diary survey methods, the proposed method is expected to reduce the frequency at which survey respondents are required to input information. On the other hand, contrary to conventional offline estimation processes, the proposed method is expected to automatically update and adapt the estimation model to current situations of the respondents themselves. Moreover, the accuracy of the method can be controlled by selecting thresholds for the confidence level.

The reminder of this article is organized as follows. The formulation of the proposed method is described in Section 2. The empirical validation is described in Section 3. Section 4 concludes this article.

2. Methodology

2.1. Overview

The proposed method infers the trip purpose, or equivalently, the type of subsequent activity. The input data for the method is data on trips from GPS mobile phones that are identified by the move-or-stay identification,[1] and limited interaction with survey respondents via the mobile phones. The method consists of two processes, namely, *estimation* and *learning*, as shown in Fig. 1a (overview in chronological order) and Fig. 2 (flowchart of the algorithm). In the estimation process, trip purposes are estimated from GPS data in almost real time (its exact timing is discussed later). According to the confidence level of the estimation, the probability of initiating the learning process is determined. It means that the system rarely proceeds to the learning process and just records the estimation result as a survey result when the confidence level is high enough. In the learning process, survey respondents are requested to respond to questions on the actual trip purpose, and the estimation model is updated by the answers. If the learning process was not initiated, the estimation result will be recorded as a survey result. Because of this online learning during the survey, the estimation model is expected to become accurate, capture the respondents' characteristics, and reduce the frequency of questions as the survey progresses. Compared to the conventional probe person survey method (e.g., Hato, 2006; Asakura et al., 2014), which automatically records trajectory of a respondent by a mobile GPS device and requests a respondent to input a trip start/end during trips and all the information about trip contexts at the inputting process of the travel diary (Fig. 1b), the proposed method is apt to require manual inputs only when the confidence level of the estimation is low.

Several options for the timing of the trip purpose estimation and questioning can be considered. The information available to the estimation and the burden on respondents varies depending on the timing. Although any timing is possible in theory, the following three are considerable in practice:

(a) start of the trip (of which the proposed method estimates the purpose),
(b) end of the trip,
(c) start of the next trip (i.e., end of the activity of which the proposed method estimates the type).

They are illustrated in Fig. 3. Clearly, the latter options can exploit more information. Practical issues regarding this timing are discussed in Section 2.5.

Fundamental notation is introduced in Section 2.2. The formulation of the proposed method is described in Section 2.3 (estimation process) and Section 2.4 (learning process). Properties of the proposed method and practical matters are discussed in Section 2.5.

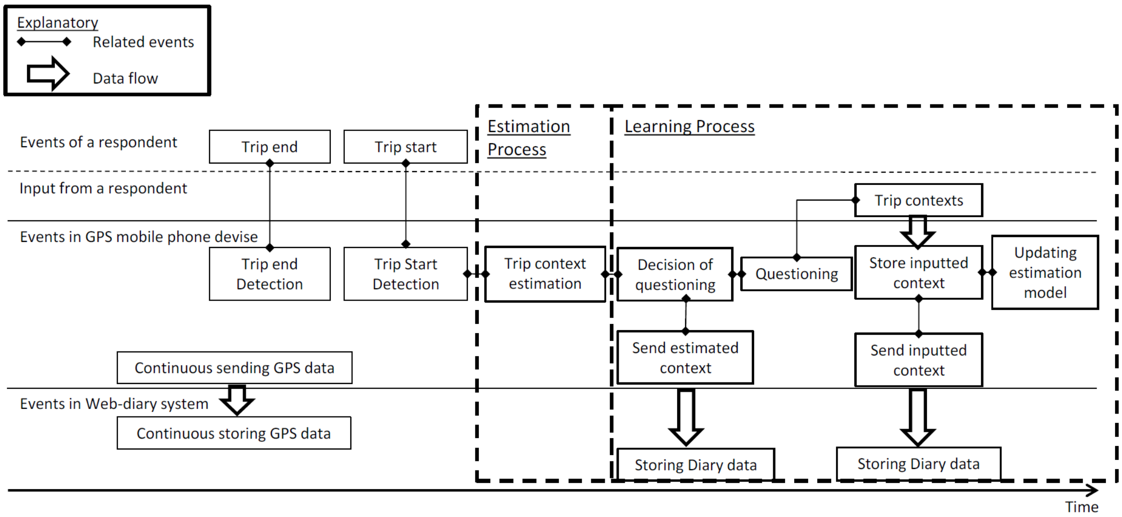
2.2. Definitions

Let $C = \{c\}$ be a set of trip purposes, $K = \{k\}$ be a set of observable trip attributes, $X^k = \{x^k\}$ be a set of possible values of a certain attribute k, and $Y = \{x^k \mid \forall k \in K\}$ be the trip situation. Historical data is denoted by $R_{n,t} = \{(c^i, Y^i) \mid \forall i \in I_{n,t}\}$, where $R_{n,t}$ is a set of data for traveler n at time step t, $I_{n,t}$ is a set of indices of trips in $R_{n,t}$, c^i indicates purpose of ith trip, and Y^i indicates situation of ith trip. See Appendix A for explanation of these definitions with a specified example.

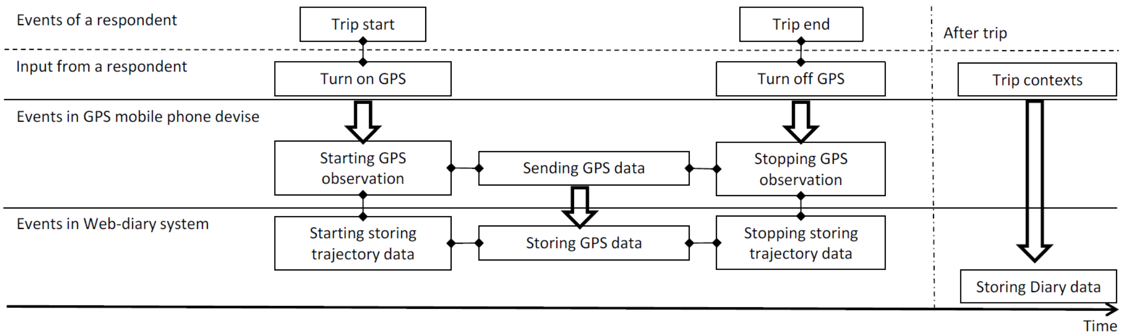
A result of the proposed method for each trip can be labeled as following states.

Correct (C): The system estimated the trip purpose correctly.
Incorrect (I): The system estimated the trip purpose incorrectly.
Question (Q): The system asked a question about the trip purpose, regardless of whether the estimation was correct or incorrect.
Success (S): The system estimated the trip purpose correctly, and did not ask a question.
Failure (F): The system estimated the trip purpose incorrectly, and did not ask a question.
True (T): The system obtained actual trip purpose as a final result, regardless of whether question was asked or not.

[1] The move-or-stay identification process automatically identifies points and times of start/end of trips based only on GPS traces (e.g., Asakura and Hato, 2004; Schüssler and Axhausen, 2009).



(a) Proposed probe person survey system with interactive online machine learning, when the timing of estimation and questioning is the start of the next trip



(b) Conventional probe person survey system

Figure 1: Relation between a respondent and a system in the proposed method, and a conventional probe person survey as a reference.

The corresponding composition ratios of the states are denoted as p_{C}, p_{I}, p_{Q}, p_{S}, p_{F}, and p_{T}, respectively. Some of these states are realized at the same time. Possible combination of states are: correct–success–true, correct–question–true, incorrect–question–true, and incorrect–failure. By the definitions, $p_{\mathrm{C}} + p_{\mathrm{I}} = 1$, $p_{\mathrm{Q}} + p_{\mathrm{S}} + p_{\mathrm{F}} = 1$, and $p_{\mathrm{T}} = p_{\mathrm{Q}} + p_{\mathrm{S}}$ hold true. The accuracy of the method is measured by p_{T}, while the burdenless-ness is measured by $1 - p_{\mathrm{Q}}$.

2.3. *Estimation process*

In general, a trip purpose estimation problem can be described as

$$\max_{c} P(c|Y), \tag{1}$$

where $P(c|Y)$ represents the probability of occurrence of trip purpose c when the observed trip situation is Y. Based on Bayes' theorem, $P(c|Y)$ can be expressed as

$$P(c|Y) = \frac{P(Y|c)P(c)}{P(Y)}, \tag{2}$$

where $P(c)$ is the probability of occurrence of trip purpose c, and $P(Y)$ is the probability occurrence of trip situation Y.

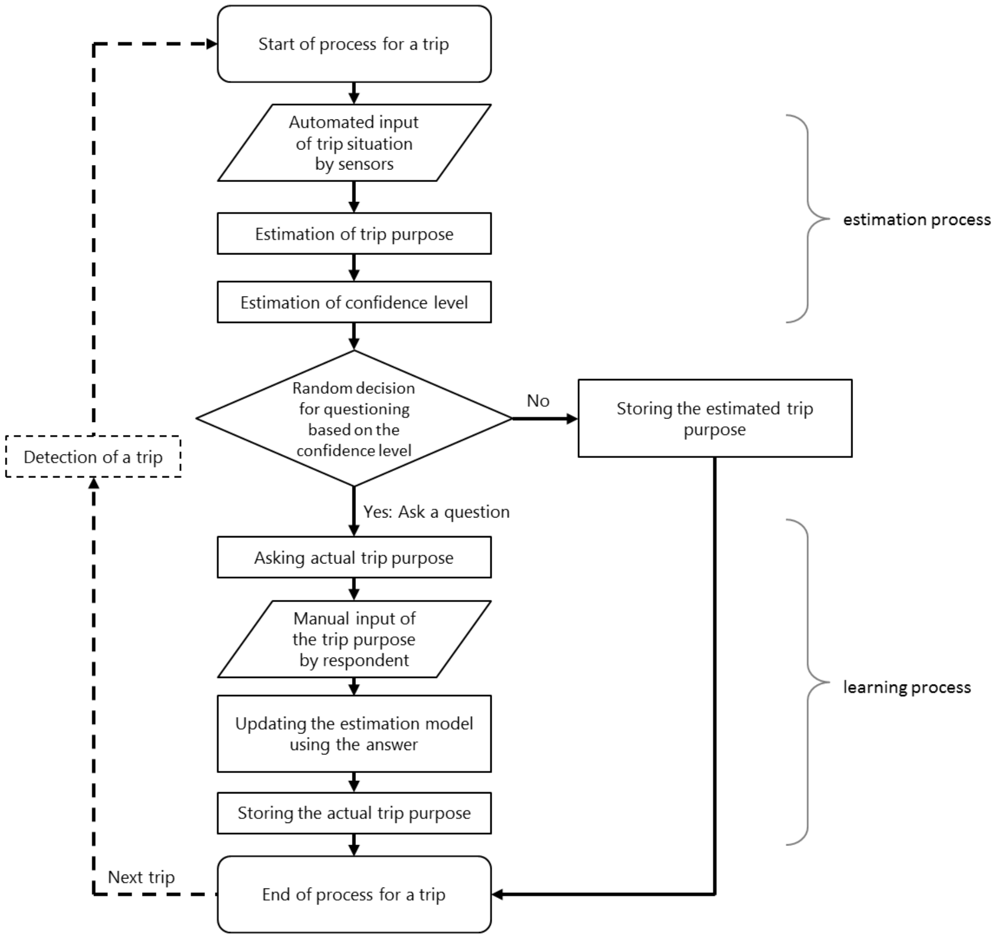


Figure 2: Process of proposed method.

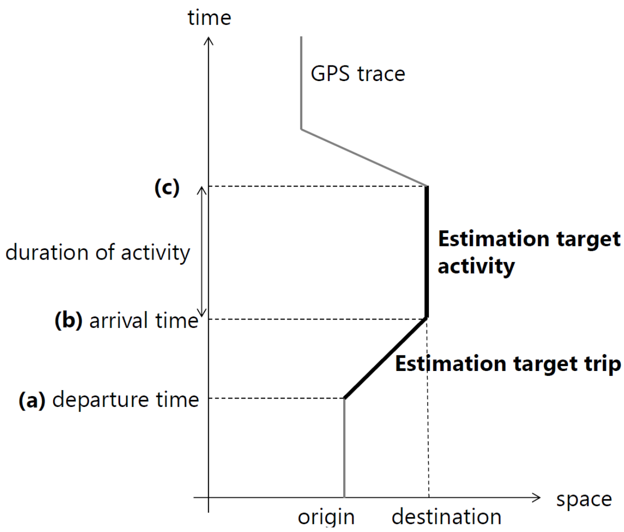


Figure 3: Timing of questioning and target of the estimation.

By employing the naïve Bayes assumption (c.f., Rish, 2001), namely, conditional independence among $P(x^k|c)$, Eq. (2) can be reduced into

$$P(c|Y) = \frac{1}{P(Y)} \prod_{k \in K} P(x^k|c)P(c). \tag{3}$$

Therefore, the trip purpose estimation problem in this article can be represented as

$$\hat{c} = \operatorname*{argmax}_{c \in C} \prod_{k \in K} P(x^k|c)P(c), \tag{4}$$

where $\hat{c}$ is an estimated trip purpose.

If a historical data $R_{n,t}$ is available, $P(c)$ and $P(x^k|c)$ for respondent n at time step t can be estimated as

$$P(c) = \frac{\sum_{i \in I_{n,t}} \delta(c, c^i)}{|I_{n,t}|} \tag{5}$$

$$P(x^k|c) = \frac{\sum_{i \in I_{n,t}} \gamma(c, x^k, c^i, Y^i)}{\sum_{z^k \in X^k} \sum_{i \in I_{n,t}} \gamma(c, z^k, c^i, Y^i)} \tag{6}$$

with

$$\delta(c, c^i) = \begin{cases} 1, & \text{if } c = c^i \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

$$\gamma(c, x^k, c^i, Y^i) = \begin{cases} 1, & \text{if } c = c^i \text{ and } x^k \in Y^i \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

by simply maximizing their likelihoods. Note that $R_{n,t}$ is a respondent-specific set, meaning that the probability functions are calculated for each individual respondent. If $R_{n,t}$ is empty (e.g., initial stage of the survey), uniform distributions can be assumed for the probability functions in order to represent no a priori information condition as in usual Bayesian estimation. The way to collect historical data $R_{n,t}$ during a survey is presented in the next section.

2.4. Learning process

Survey respondents are occasionally asked their actual trip purpose based on the confidence level of the estimation result; and then the historical data $R_{n,t}$ is updated by the answer (i.e., the answer is added to the historical data). Specifically, a trip purpose question randomly appears as an estimation failure rate of each estimation is expected to be (approximately) identical to an acceptable failure rate, p_{af}, which is pre-given by the analysts in the planning phase of the survey. The confidence level is $P(\hat{c}|Y)$ and determined by

$$P(\hat{c}|Y) = \frac{\prod_{k \in K} P(x^k|\hat{c})P(\hat{c})}{\sum_{c \in C} \prod_{k \in K} P(x^k|c)P(c)}. \tag{9}$$

The probability of a question appearing, p_{aq}, is defined as

$$p_{\text{aq}} = \max\left\{0,\ 1 - \frac{p_{\text{af}}}{1 - P(\hat{c}|Y)}\right\}. \tag{10}$$

When the question does not appear or is not answered by a respondent, the estimation result described in Eq. (4) is recorded, and the probability functions are not updated. When respondents answer the question, the probability functions are updated using the answer and Eqs. (5) and (6).

2.5. *Discussion*

By applying the proposed method to the same respondent for a long-term period, the proposed method is expected to learn the activity pattern of the respondent. At the initial stage of the survey, the survey system will ask questions almost every time (i.e., high p_{Q}). As the survey progresses, the system will then learn the characteristics of trips (i.e., estimate $P(x^k|c)$ accurately by re-calculating Eqs. (5) and (6) with updated $R_{n,t}$) based on interaction with the respondent, and infer the purposes automatically (i.e., low p_{Q} and high p_{S}). The learning speed would differ depending on the personal activity pattern—the higher the regularity of trips is (e.g., commuting), the faster the learning speed is. Meanwhile, the error of the system is controlled by the acceptable failure rate p_{af}, so that the quality of the results is guaranteed. In cases where irregular trips or sudden behavioral changes are observed, the system will start asking questions again and eventually adapt to the new patterns.

In terms of the survey accuracy, high true rate p_{T} is preferable. On the other hand, in terms of the survey burden, low question rate p_{Q} is preferable. Meanwhile, failure rate $p_{\mathrm{F}} = 1 - p_{\mathrm{T}}$ is expected to be approximately identical to the acceptable failure rate p_{af}, which is pre-given by the survey designer. In addition, the higher p_{af} is given, the lower p_{Q} is expected according to Eq. (10). Consequently, we can expect that there is a trade-off relation between accuracy (high p_{T}) and burdenless-ness (low p_{Q}) of the survey, and this can be controlled by the value of the acceptable failure rate p_{af}.

Several timings for questioning can be considered, as mentioned in Section 2.1. Each of them has different practical advantages and disadvantages. In terms of the observable information, it can be expected that the timing (c) will be the most accurate, followed by the timing (b). This is because they can utilize information on the destination of trips and/or duration of activity from which purpose/type is estimated. Regarding the burden on respondents, the timings (a) and (c) are possibly not preferable for respondents who are driving, while (b) may not be suitable for respondents who are performing busy activities. If a respondent missed a question, the estimated purpose can be substituted as the survey result without waiting for an answer. The accuracy of the method might be lowered from the expected value in this case.

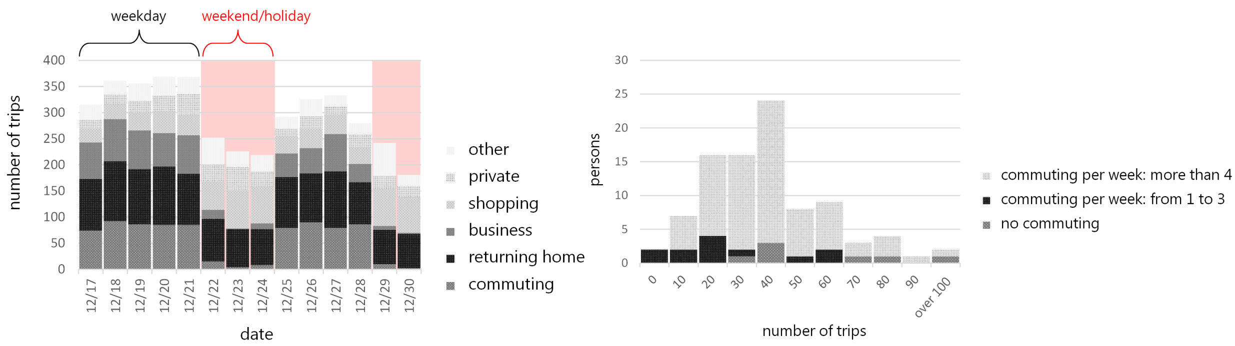
These properties of the proposed method, namely, learning speed, effect of acceptable failure rate, and timing for questioning, are empirically investigated in Section 3.

The proposed method uses a naïve Bayes classifier (Rish, 2001), which is not always accurate compared to other advanced methods, such as discrete choice models (Oliveira et al., 2014), Bayesian networks, and neural networks (Feng and Timmermans, 2016). However, naïve Bayes classifiers require smaller amounts of sample data, compared with such advanced methods. This feature is essentially beneficial for the proposed approach; because the expected size of sample data for the proposed approach is limited. For example, the size of sample data for a usual respondent (who makes three trips per day on average) is expected to be only 90 trips per month at the maximum. In fact, the sample size must be substantially smaller than 90 in this case, since automatically inferred purposes are not considered as samples. Such small sample sizes would not be sufficient to train or calibrate the aforementioned advanced methods. Besides, the naïve Bayes classifier is computationally efficient. This is also beneficial to the proposed approach, which trains and updates its classifier iteratively as the survey progresses. Therefore, its implementation on mobile phones is remarkably easy and battery-friendly.

The proposed method can be considered as a generalization of certain existing activity-travel survey methods in the following sense. In the case of $p_{\mathrm{af}} = 0$, the proposed method is identical to conventional travel-diary-based surveys assisted by GPS, as respondents manually input all the trip purposes. In the case of $p_{\mathrm{af}} = 1$ and when intentionally calibrated initial distributions for $P(c)$ and $P(x^k|c)$ are given, the proposed method can be considered as a trip purpose inference method with a predetermined inference model, as all the trip purposes are estimated without manual inputs or online learning.

3. Empirical Analysis

The proposed method is examined by simulating it on actual behavioral data obtained from a past probe person survey.



(a) trip purposes over the days

(b) total number of trips of each respondent, characterized by average number of commuting trips per week

Figure 4: Basic properties of the survey data.

3.1. Validation method

3.1.1. Data

The data for validation is collected from a probe person survey using web-based diary and mobile communication systems, conducted by Ministry of Land, Infrastructure, Transport and Tourism and Matsuyama city. The data include trip purpose, origin/destination places, and beginning/end of trips, manually inputted by the respondents. The diary data were collected in Matsuyama city in Japan from December 17 to December 30, 2007. The respondents consist of usual workers and their families. The number of respondents is 92. The respondents made 4120 trips during the period. This corresponds to 44.8 trips/person and 3.2 trips/person/day. The basic properties of the data, namely, the trip purposes over the days, and the total number of trips of each respondent, are shown in Fig. 4. The days with pink-colored background in Fig. 4a, namely, 12/22, 12/23, 12/24, 12/29, and 12/30, are weekends and holidays.

The trip attributes in the data are employed for the trip purpose estimation and learning processes of the proposed method. In the process described in Section 2.4, when a question arises, the trip purpose data are regarded as the answer to the arisen question, and are used for the learning process. When a question does not arise, the trip purpose data are only used for validating whether an estimation result is correct or not. Since the method asks the questions randomly, Monte Carlo simulation is conducted in this validation.

3.1.2. Specification of the model and scenario parameters

In this validation, the purpose of each trip is defined as context c. They are defined as

$$C = \{\text{commuting}, \text{returning home}, \text{business}, \text{shopping}, \text{private}, \text{others}\}. \tag{11}$$

The business purposes represent the trips where travelers travel between workplaces and other places such as their clients' offices.

Trip attributes available to the proposed method vary depending on the timing of the estimation as discussed in Section 2.5. The three timings of estimation and questioning discussed in Section 2.1 are considered, namely (a) start of the trip (of which the purpose is estimated), (b) end of the trip, and (c) start of the next trip. The trip attributes set corresponding to the timing are defined as

$$K_\mathrm{a} = \{\text{day of week}, \text{time of day at departure}, \text{location of origin}\}, \tag{12}$$
$$K_\mathrm{b} = \{\text{day of week}, \text{time of day at arrival}, \text{location of destination}\}, \tag{13}$$
$$K_\mathrm{c} = \{\text{day of week}, \text{time of day at arrival}, \text{location of destination}, \text{duration of activity}\}, \tag{14}$$

respectively. Note that departure and arrival information should not be simultaneously included into an attributes set, as they are often strongly correlated and thus violate the underlying assumption of the naïve Bayes method. Each variable is discretized to adapt to the employed naïve Bayes method. Day of week is defined as $X^{\text{day of week}} = \{\text{weekday}, \text{weekend-or-holiday}\}$. Places $X^{\text{location of origin}}$ and $X^{\text{location of destination}}$ are defined as a square-shaped mesh discretized with 100 m times 100 m rectangle. Time of day $X^{\text{time of day at departure}}$ and $X^{\text{time of day at arrival}}$ is discretized into three-hour-length bins, and that of $X^{\text{duration of activity}}$ is one hour. Durations larger than 12 hours are classified in the bin of 12 hours.

The other important parameter of the proposed method is the acceptable failure rate, p_{af}. For this, the following four parameter values are considered: 1%, 5%, 10%, and 15%.

In the following analyses, the scenario with timing (b) and $p_{\text{af}} = 5\%$ is considered as a *reference scenario* because of its moderateness.

3.2. Results

The overall results and comparison among scenarios (i.e., timing of questioning and acceptable failure rate) are presented in Section 3.2.1. Detailed analyses on the reference scenario are presented in Sections 3.2.2 and 3.2.3. Finally, their implications are discussed in Section 3.3.

3.2.1. Overall performance, acceptable failure rate, and timing of questioning

The accuracy of the proposed method with question timing (b) is shown in Fig. 5 as a time-series. The horizontal axis represents the number of days from the survey beginning, while the vertical axis represents the composition rate of questions, success, and failure rates as defined in Section 2.5, averaged over all the respondents. The percentile values indicate proportions in Monte Carlo simulation replications. The days with pink-colored background are weekends and holidays.

In the reference scenario with $p_{\text{af}} = 5\%$ shown in Fig. 5b, the question arose in more than 90% of trips in the first day of the survey. As the survey progressed in the weekdays, the question rate p_{Q} decreased monotonically, the success rate p_{S} increased monotonically, and the failure rate p_{F} remained almost constant. The performance of the last weekday (i.e., December 28th) was $p_{\text{Q}} = 63\%$, $p_{\text{S}} = 34\%$, and $p_{\text{F}} = 3\%$. In the weekends, similar tendencies were observed, but the question rate was higher than that in the weekdays. On the other hand, the failure rate remained almost constant at 6% throughout the survey period. Moreover, the results of different simulation replications show a narrow distribution (i.e., difference between the 5%tile value and the 95%tile value is small), implying that the method is fairly stable.

In the cases of different p_{af} values, similar tendencies were observed as well (Fig. 5a, 5c, 5d). On the other hand, the success rate p_{S} showed a tendency to increase as p_{af} increases, and the trade-off relation between p_{T} and $1 - p_{\text{Q}}$ discussed in Section 2.5 was clearly observed. However, the values of the average failure rates were not identical to the pre-given p_{af} values. This might be due to the conditional independence assumption in the naïve Bayes model, not being exactly satisfied.

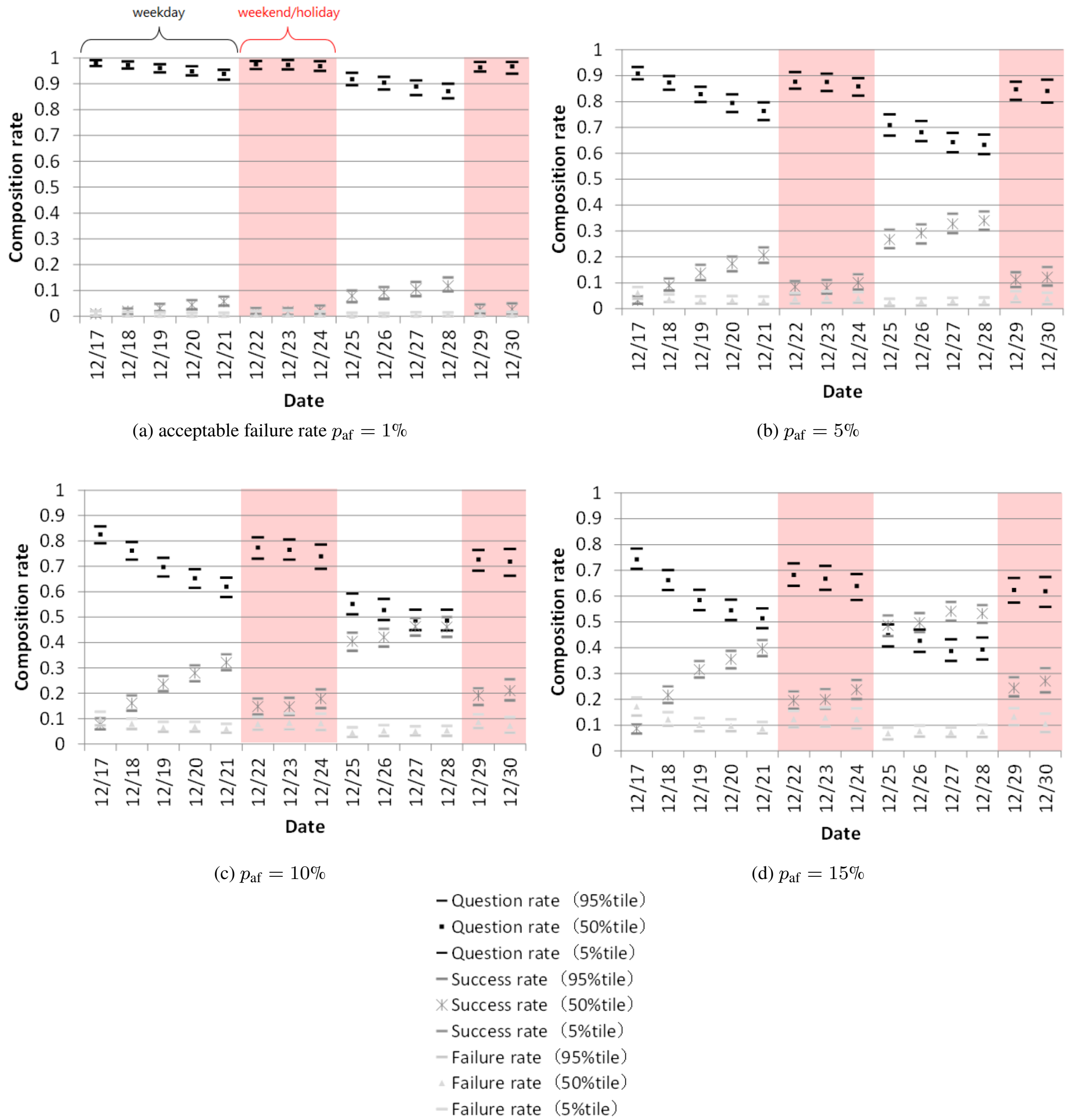
The average performance, namely, average true and question rates of every scenario, averaged over the entire survey period, is shown in Tabs. 1 and 2. From Tab. 2, it was confirmed that p_{T} is roughly equal to $1 - p_{\text{af}}$ in most of the scenarios. For the scenarios with the same timing of questioning, the aforementioned trade-off relation between p_{T} and $1 - p_{\text{Q}}$ can be found again; namely, p_{Q} decreases as p_{af} ($\simeq 1 - p_{\text{T}}$) increases in Tab. 2. Regarding the timing of questioning, it was suggested that timing (c) is preferable in terms of accuracy; while timing (b) is less preferable but better than (a).

3.2.2. Trip purpose

Average performance regarding actual trip purposes in the reference scenario (i.e., $p_{\text{af}} = 5\%$ and timing (b)) is shown in Tab. 3. The commuting trips had the lowest p_{Q} among the purposes, because they usually have strong regularity in time and destinations. The p_{Q} in shopping, private, and others were higher than those of the other trip purposes. This is because activities in these trip purposes did not have strong regularity. Regarding p_{T}, all of the purposes showed similar values, close to the pre-given values for p_{af} as expected. However, p_{T} in the shopping, private, and others were slightly lower than p_{af}. This might be because these three trips were not easily distinguishable from one another due to the lack of regularity.

3.2.3. Individual characteristics

The capability of the proposed method to capture traveler heterogeneity is investigated here. The average performance for each respondent in the reference scenario is shown in Fig. 6 where each cross marker indicates



(a) acceptable failure rate $p_{\text{af}} = 1\%$

(b) $p_{\text{af}} = 5\%$

(c) $p_{\text{af}} = 10\%$

(d) $p_{\text{af}} = 15\%$

Figure 5: Estimation results in time-series with timing (b).

the average performance of a respondent. It can be observed that while p_{Q} varies among respondents, p_{T} is almost the same for all respondents. This means that, although the reduction of survey burden varies among respondents due to traveler heterogeneity, the method eventually obtained accurate results by questioning.

A more dynamical aspect of the proposed method, namely, the learning speed of the method for each respondent is shown in Fig. 7. The x–y plot shows that, if the learning process (i.e., questioning and learning) of the proposed method were terminated on the xth day, $y \times 100\%$ of the whole trips of each respondent can be successfully estimated.[2] The percentile indicates the composition of respondents. For example, if the learning process were terminated on the

[2]This value is estimated for each respondent in every day in every Monte Carlo simulation. Each value is derived from 100 trips obtained by

Table 1: Average true rate p_{T} in each scenario.

Timing	Acceptable failure rate p_{af}			
	1%	5%	10%	15%
(a)	98.8%	94.4%	89.3%	84.5%
(b)	99.2%	96.3%	92.9%	89.2%
(c)	99.0%	95.6%	91.7%	87.9%

Table 2: Average question rate p_{Q} in each scenario.

Timing	Acceptable failure rate p_{af}			
	1%	5%	10%	15%
(a)	95.3%	81.8%	69.2%	58.9%
(b)	94.5%	79.3%	66.3%	56.3%
(c)	90.2%	72.2%	59.2%	49.7%

Table 3: Average performance regarding trip purposes in the reference scenario.

Trip purpose	True rate p_{T}	Question rate p_{Q}
Commuting	98.2%	68.8%
Returning home	98.1%	75.6%
Business	97.4%	77.2%
Shopping	94.2%	88.3%
Private	92.1%	90.0%
Other	93.4%	87.4%

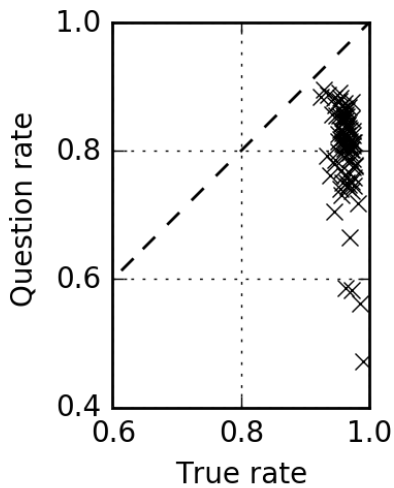


Figure 6: Average performance regarding each respondent in the reference scenario.

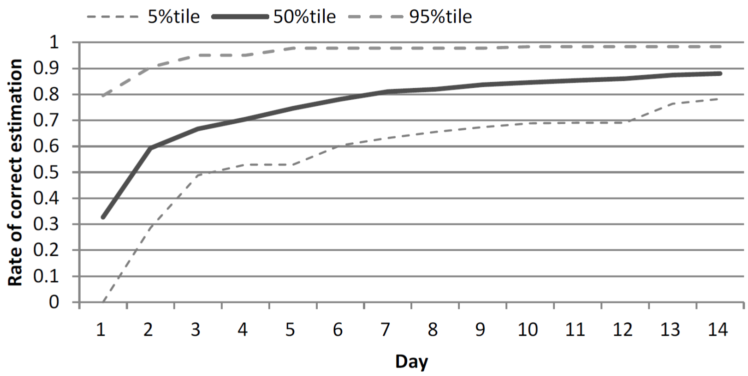


Figure 7: Learning speed over respondents in the reference scenario.

7th day, purposes of more than 82% of trips of half (i.e., 50%tile) of respondents could be successfully estimated. Similarly, those of more than 64% of trips of 95% of respondents and those of more than 98% of trips of 5% of respondents could be successfully estimated in this case.

According to Fig. 7, the accuracy of the estimation model increased monotonically as the survey progresses, regardless of respondents. Meanwhile, the learning states of the proposed method significantly varied among respondents, especially at the initial stage of the survey. This is because of their heterogeneity in travel regularity. However, the difference also decreased almost monotonically as the survey progresses. Note that the respondents with fast learning speed in Fig. 7 correspond to those with low p_{Q} in Fig. 6.

3.3. Discussion

It was confirmed that the performance of the proposed method in the weekdays improved monotonically as the survey progresses, regardless of respondents. On the other hand, the failure rate also remained almost constant regardless of respondents. Moreover, while the values of the question rate p_{Q} varied among trip purposes and respondents depending on their regularity, the true rate p_{T} was almost the same among them. These results suggest that the method properly captured heterogeneity among respondents, by dynamically updating the inference model for each respondent. This feature might be enabled by the naïve Bayes classifier, which only requires a small sample size. However, it was also confirmed that the simplification in the naïve Bayes classifier introduced slight systematic errors, for instance on failure rates in private and shopping trips.

Increasing the acceptable failure rate greatly reduced survey burden, at the cost of data quality. Therefore, the optimal value of the acceptable failure rate depends on the requirements of a survey, and should be selected carefully by the survey planner.

Regarding the timing of questioning, timing (c) was revealed to be the most-efficient in terms of the estimation performance, followed by timing (b). Timing (a) was relatively inefficient. These results are reasonable as discussed in Section 2.5. Therefore, in a practical survey, it would be considerable to set the timing of questioning to either (b) or (c), depending on respondents' situation (e.g., driving or not, activity busyness).

4. Conclusions

This article has proposed a framework for an interactive activity-travel survey method for semi-automated trip purpose inference. The proposed method has been designed to dynamically update the inference model by asking a respondent, when the confidence level of the estimation results are not high enough. As a result, the proposed method would be able to accurately infer the trip purposes by capturing traveler heterogeneity, while being independent of preceding data collection and calibration.

randomly sampling from trips made by each respondents.

The proposed method was examined by simulation based on actual behavioral data collected from a conventional probe person survey. The results suggest that the frequency of inputs by survey respondents can be significantly reduced, while maintaining high accuracy of the obtained data as expected. For example, the method successfully estimated certain types of trips (e.g., commuting) and behaviors of certain respondents (e.g., those whose activity-travel behavior is recurrent) because of the learning process, and reduced survey burden on them. At the same time, although the method could not always precisely estimate some other kinds of trips, the method eventually obtained accurate results because of the questioning. Therefore, the proposed method could be useful to reduce the burden on respondents in a long-term survey, while keeping the data quality high and capturing traveler heterogeneity.

Several future research directions can be proposed. First, identifying similar types of travelers and applying the same activity inference model to them would be valuable in order to employ advanced learning methods, which are more accurate and require larger sample sizes (c.f., Section 2.5). Second, the application of the proposed approach to other non-observable trip attitudes, such as travel companions and fares, is considerable. Third, field implementation of the proposed method and a large-scale survey are now being conducted by the authors to investigate detailed characteristics of the method.

Acknowledgements

The authors would like to thank Prof. Eiji Hato (The University of Tokyo), Mr. Tomonori Ishii (Matsuyama City), and Mr. Michihiro Mori (Fukken Co., Ltd.) for their cooperation in using the Matsuyama data. Part of this work was financially supported by the Japan Society for the Promotion of Science (KAKENHI Grant-in-Aid for Scientific Research (S) 26220906).

Appendix A: Example of trip situation

For easier understanding, following explain the definition in Section 2.2 using a specified example.

Let trip purposes C be $\{c_1, c_2\}$ where c_1 and c_2 represent returning-home and commuting, respectively. Let trip attributes K be $\{k_1, k_2, k_3\}$ where k_1, k_2, and k_3 represent weekday-or-not, departure time, and location of origin, respectively. Regarding X^k, let values for weekday-or-not X^1 be $\{x^1_1, x^1_2\}$ where x^1_1 and x^1_2 indicate weekday and holiday, respectively. Also, let values for departure time X^2 be $\{x^2_1, x^2_2\}$ where x^2_1 and x^2_2 indicate a.m. and p.m., respectively. Finally, let values for the location of the origin X^3 be $\{x^3_1, x^3_2, x^3_3\}$ where x^3_1, x^3_2 and x^3_3 indicate home, office, and others, respectively.

An example of ith trip can be one with purpose $c^i = c_2$ under situation $Y^i = \{x^1_1, x^2_2, x^3_2\}$, which can be translated as "a returning-home trip in weekday afternoon, departed from office". The proposed method first observes $Y^i = \{x^1_1, x^2_2, x^3_2\}$ automatically, and then tries to infer the value of c^i in the estimation process. If the method asks a question to the survey respondent and it is answered properly, $c^i = c_2$ will be observed and $R_{n,t+1}$ will be constructed by adding element $(c_2, \{x^1_1, x^2_2, x^3_2\})$ to $R_{n,t}$ in the learning process.

References

Alexander, L., Jiang, S., Murga, M., González, M. C., 2015. Origin-destination trips by purpose and time of day inferred from mobile phone data. Transportation Research Part C: Emerging Technologies 58, 240–250.

Arentze, T. A., Timmermans, H. J. P., 2009. A need-based model of multi-day, multi-person activity generation. Transportation Research Part B: Methodological 43 (2), 251–265.

Asakura, Y., Hato, E., 2004. Tracking survey for individual travel behaviour using mobile communication instruments. Transportation Research Part C: Emerging Technologies 12 (3–4), 273–291.

Asakura, Y., Hato, E., Maruyama, T., 2014. Behavioural data collection using mobile phones. In: Rasouli, S., Timmermans, H. J. P. (Eds.), Mobile Technologies for Activity-Travel Data Collection and Analysis. IGI Global, pp. 17–35.

Auld, J., Williams, C., Mohammadian, A., Nelson, P., 2009. An automated GPS-based prompted recall survey with learning algorithms. Transportation Letters 1 (1), 59–79.

Axhausen, K. W., Löchl, M., Schlich, R., Buhl, T., Widmer, P., 2007. Fatigue in long-duration travel diaries. Transportation 34 (2), 143–160.

Bhat, C. R., Castro, M., Khan, M., 2013. A new estimation approach for the multiple discrete–continuous probit (MDCP) choice model. Transportation Research Part B: Methodological 55, 1–22.

Cottrill, C. D., Pereira, F. C., Zhao, F., Dias, I. F., Lim, H. B., Ben-Akiva, M. E., Zegras, P. C., 2013. Future Mobility Survey. Transportation Research Record: Journal of the Transportation Research Board 2354 (1), 59–67.

Draijer, G., Kalfs, N., Perdok, J., 2000. Global positioning system as data collection method for travel research. Transportation Research Record: Journal of the Transportation Research Board (1719), 147–153.

Feng, T., Timmermans, H. J. P., 2016. Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data. Transportation Planning and Technology 39 (2), 180–194.

Gong, L., Morikawa, T., Yamamoto, T., Sato, H., 2014. Deriving personal trip data from GPS data: A literature review on the existing methodologies. Procedia - Social and Behavioral Sciences 138, 557–565.

Han, G., Sohn, K., 2016. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model. Transportation Research Part B: Methodological 83, 121–135.

Hasan, S., Ukkusuri, S. V., 2014. Urban activity pattern classification using topic models from online geo-location data. Transportation Research Part C: Emerging Technologies 44, 363–381.

Hasan, S., Zhan, X., Ukkusuri, S. V., 2013. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing.

Hato, E., 2006. Development of MoALs (mobile activity loggers supported by GPS-phones) for travel behavior analysis. In: Transportation Research Board 85th Annual Meeting.

Kim, Y., Pereira, F. C., Zhao, F., Ghorpade, A., Zegras, P. C., Ben-Akiva, M., 2014. Activity recognition for a smartphone based travel survey based on cross-user history data. In: 2014 22nd International Conference on Pattern Recognition. pp. 432–437.

Kitamura, R., 1990. Panel analysis in transportation planning: An overview. Transportation Research Part A: General 24 (6), 401–415.

Kitamura, R., Yamamoto, T., Fujii, S., 2003. The effectiveness of panels in detecting changes in discrete travel behavior. Transportation Research Part B: Methodological 37 (2), 191–206.

Kusakabe, T., Asakura, Y., 2014. Behavioural data mining of transit smart card data: A data fusion approach. Transportation Research Part C: Emerging Technologies 46, 179–191.

Kusakabe, T., Tsubota, T., Bhaskar, A., 2016. Validation study of naïve Bayes probabilistic model for transit passengers' trip purpose estimation: Case study exploiting detailed Brisbane household travel survey data. In: Transportation Research Board 95th Annual Meeting.

Murakami, E., Wagner, D. P., 1999. Can using global positioning system (GPS) improve trip reporting? Transportation Research Part C: Emerging Technologies 7 (2-3), 149–165.

Oliveira, M., Vovsha, P., Wolf, J., Mitchell, M., 2014. Evaluation of two methods for identifying trip purpose in GPS-based household travel surveys. Transportation Research Record: Journal of the Transportation Research Board (2405), 33–41.

Prelipcean, A. C., Susilo, Y. O., Gidofalvi, G., 2017. Collecting travel diaries: Current state of the art, best practices, and future research directions. In: Proceedings of the 11th International Conference on Transport Survey Methods.

Rasouli, S., Timmermans, H. J. P. (Eds.), 2014. Mobile technologies for activity-travel data collection and analysis. IGI Global.

Rish, I., 2001. An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. Vol. 3. pp. 41–46.

Schüssler, N., Axhausen, K. W., 2009. Processing raw data from global positioning systems without additional information. Transportation Research Record: Journal of the Transportation Research Board (2105), 28–36.

Shen, L., Stopher, P. R., 2013. A process for trip purpose imputation from Global Positioning System data. Transportation Research Part C: Emerging Technologies 36, 261–267.

Shen, L., Stopher, P. R., 2014. Review of GPS travel survey and GPS data-processing methods. Transport Reviews 34 (3), 316–334.

Siripirote, T., Sumalee, A., Ho, H. W., Lam, W. H. K., 2015. Statistical approach for activity-based model calibration based on plate scanning and traffic counts data. Transportation Research Part B: Methodological 78, 280–300.

Washington, S., Ravulaparthy, S., Rose, J. M., Hensher, D., Pendyala, R., 2014. Bayesian imputation of non-chosen attribute values in revealed preference surveys. Journal of Advanced Transportation 48 (1), 48–65.

Wolf, J., Guensler, R., Bachman, W., 2001. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. Transportation Research Record: Journal of the Transportation Research Board (1768), 125–134.

Xiao, G., Juan, Z., Zhang, C., 2016. Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. Transportation Research Part C: Emerging Technologies 71, 447–463.