Automated and Adaptive Activity-Travel Survey using Online Interaction with Travelers

Toru SEO

Postdoctoral Researcher, Asakura Lab., Tokyo Institute of Technology

Joint work with Dr. Kusakabe and Prof. Asakura

The 7th International BinN Seminar @ UTokyo

20160925

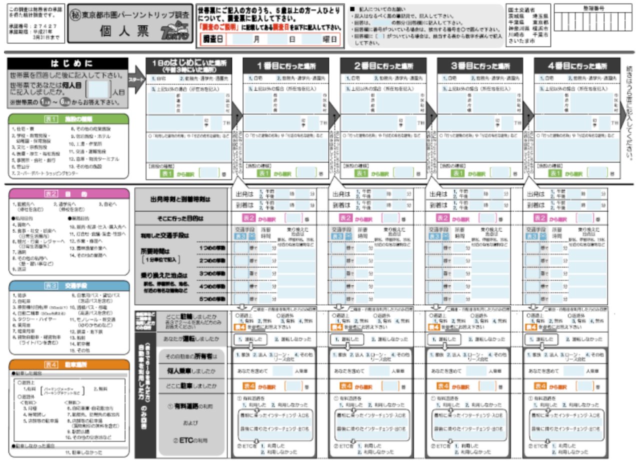
Travel behavior survey

- Travel behavior data is essential for activity/travel behavior modeling

traveler id	trip id	origin	destination	mode	departure time	⋯
1	1	home	office	train	8:30	
1	2	office	shop	walk	17:00	
2	1	home	restaurant	car	11:30	
⋮						

- Actual activity/travel data is often collected by conducting a survey
  – realistic compared to stated preference (virtual) data

Conventional survey methods



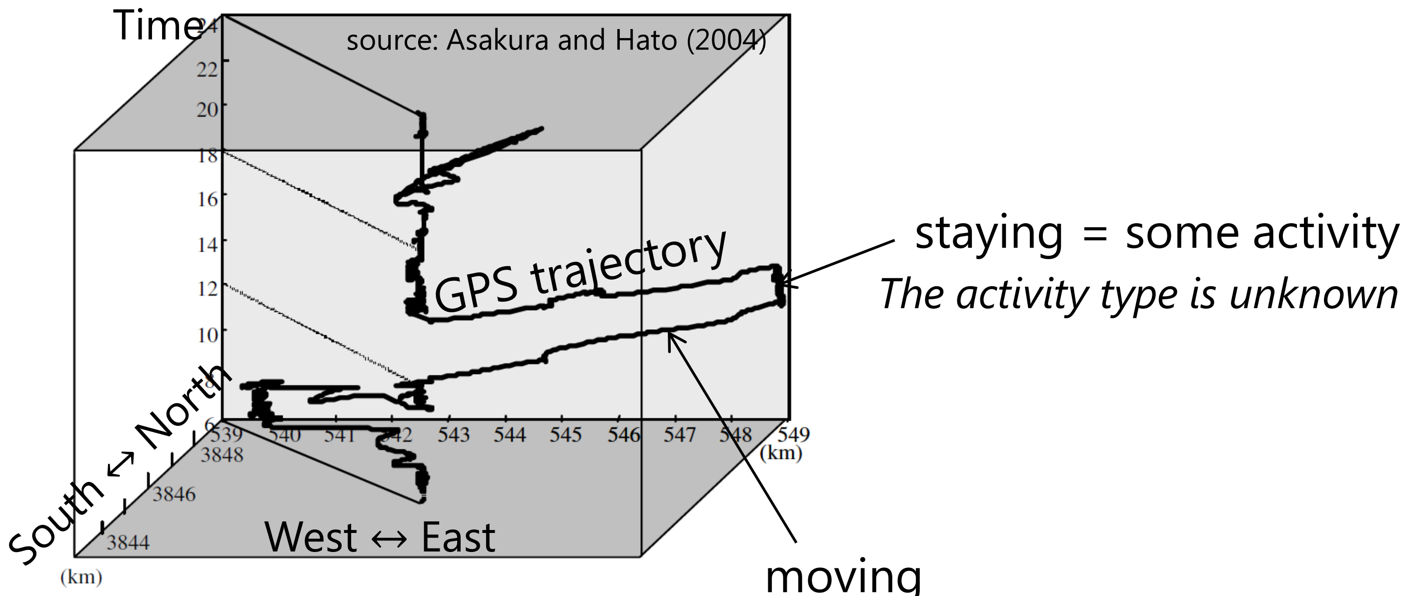
Questionary survey

- Travel data is collected by paper- or web-based questionary

- Limitations:
  - burden for survey participants: *where did you go? when? how? why? with whom? etc.*
  - inaccurate due to incomplete memory

  →Long-term, large-scale, and accurate data collection is difficult (panel attrition, fatigue)

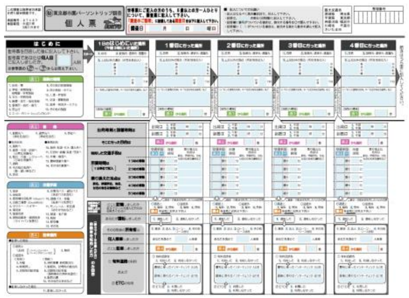
Conventional survey methods



GPS data collection

- accurate spatiotemporal data
- Activity type (trip purpose) is missing

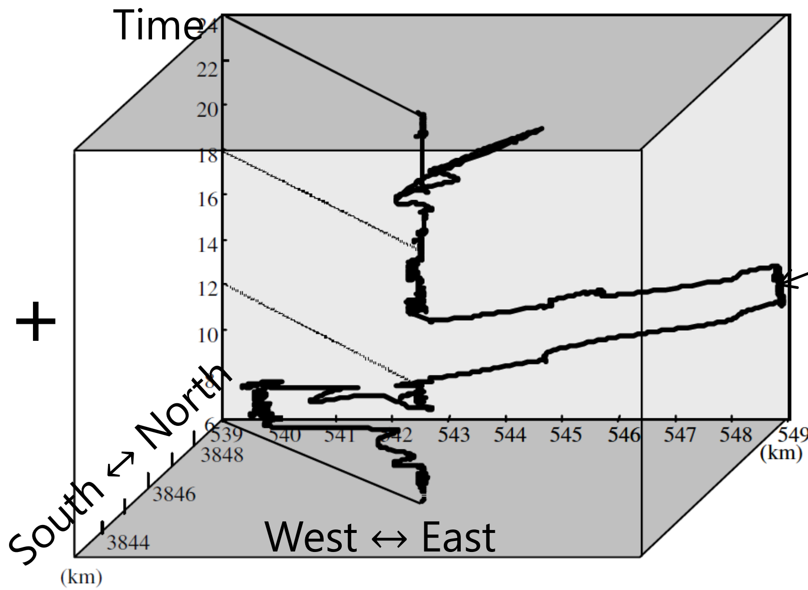
Conventional survey methods



survey sheet

+

web/app interface



staying

The activity type is asked by questionary survey

GPS data collection + questionary survey

- accurate spatiotemporal data
- relatively accurate activity data
  - GPS log can help memory recalling

- burden for participants
  - Extensive manual input is still mandatory

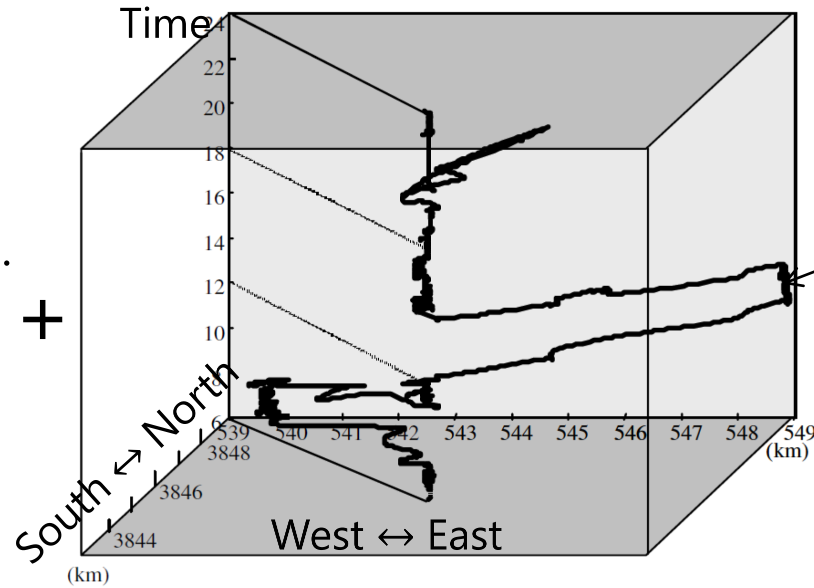
Existing advanced survey methods

land use/facility info.

$$\frac{e^{V_i}}{\sum_j e^{V_j}}$$

calibrated model

+



The place is shop
↓
The activity must be shopping

staying

Existing behavior model predicts that he will go shopping
↓
It must be shopping

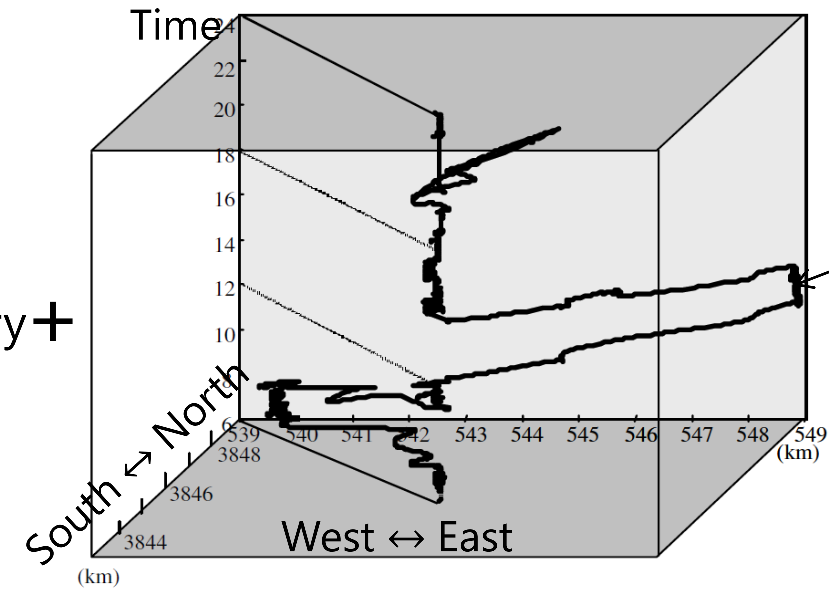
GPS data collection + imputation based on *a priori* info.

- Automatic imputation based on a priori information (eg: Wolf et al., 2001; Shen and Stopher, 2013; Gong et al., 2013)
  - land use
  - behavior model calibrated by using existing data

- Collected data may have several limitation
  - Traveler heterogeneity is ignored
  - The data is not suitable for behavior modeling purposes

Existing advanced survey methods

calibrated model
based on preliminary +
survey



staying

During the preliminary survey,
he went there for shopping
↓
It must be shopping

GPS data collection + imputation based on *offline* info.

(Kim, Ben-Akiva, et al., 2014, 2015)

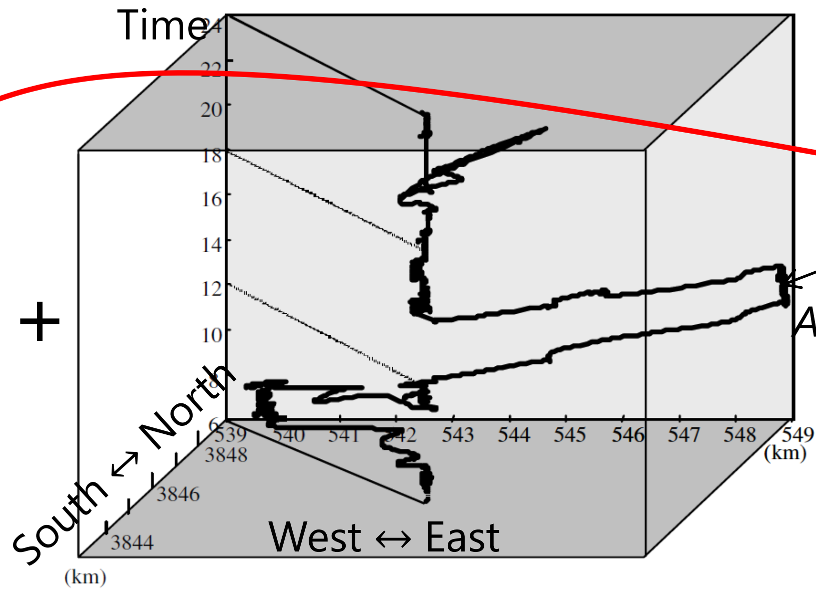
- Automatic imputation based on preliminary survey
  - Data is collected by preliminary survey for the same participants
  - Traveler heterogeneity can be captured

Proposed survey method

online update

model

+



If there is no clue, the system asks a question
↓
answer: shopping

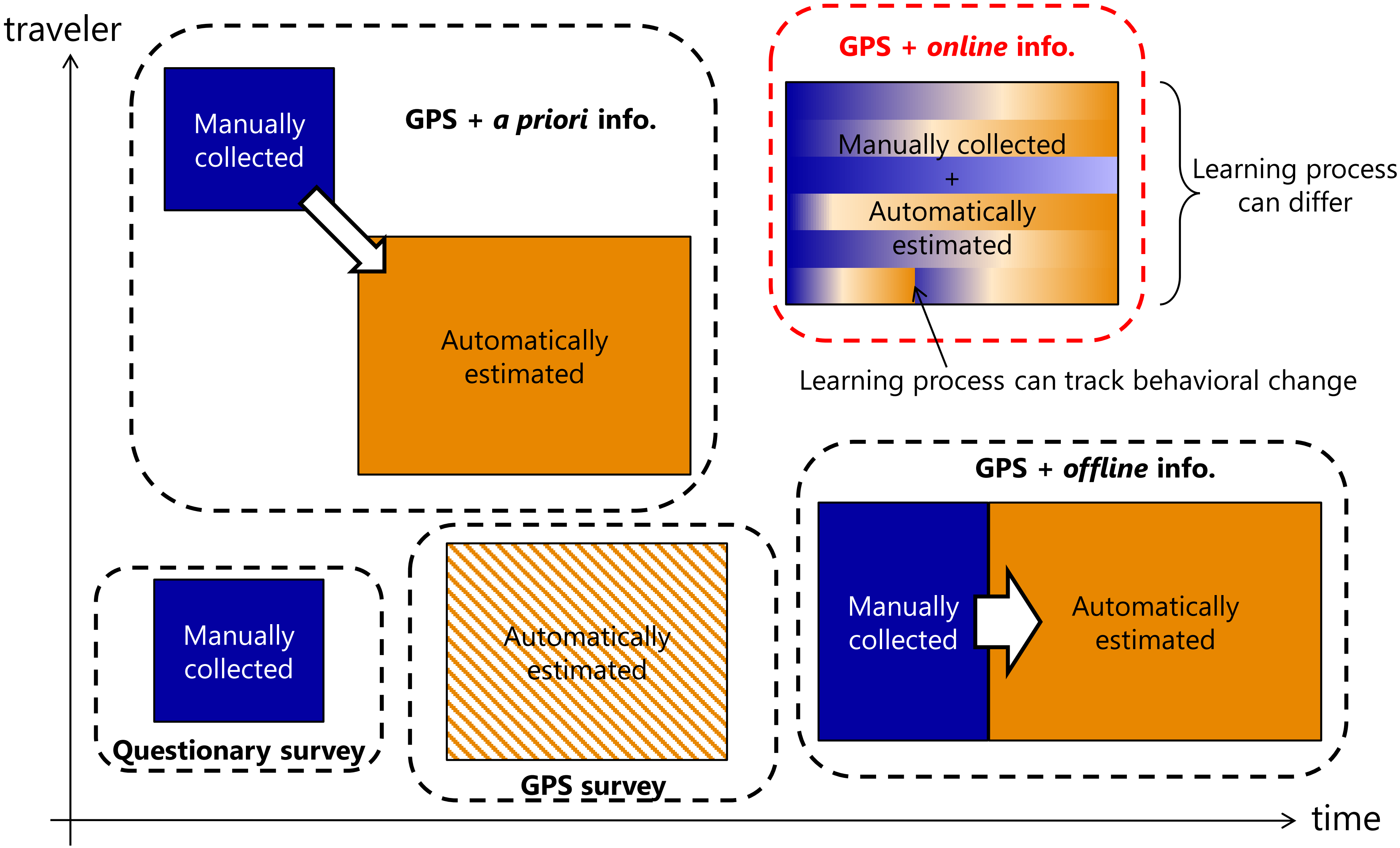
According to the up-to-date model, it seems shopping
↓
I must be shopping

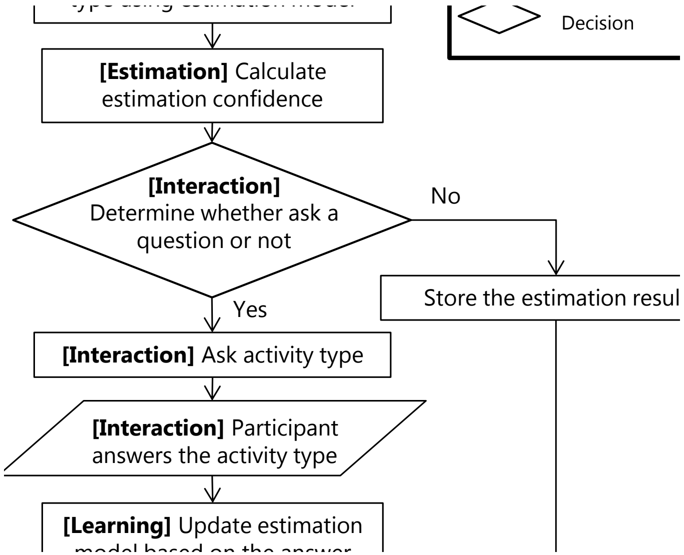
GPS data collection + imputation based on *online* info. (Kusakabe et al., 2015; Seo et al., 2016)

- Automatic imputation based on data collected from previous online interaction
  - online interaction: the survey system will ask a question automatically and dynamically

- Possible features of the proposed method:
  - Reducing frequency of questions
  - Keeping quality of data high
  - Considering traveler heterogeneity
  - Adaptive tracking of behavioral change
  - Automatic process (No need of manual control by survey administrator)

Illustration of methods





$$\hat{c} = \mathrm{argmax}_c\, P(c|Y)$$

$$\hat{c} = \mathrm{argmax}_c \prod_k P(x_k|c)P(c)$$

$$\{(c_1, Y_1), (c_2, Y_2), \dots (c_i, Y_i), \dots\}$$

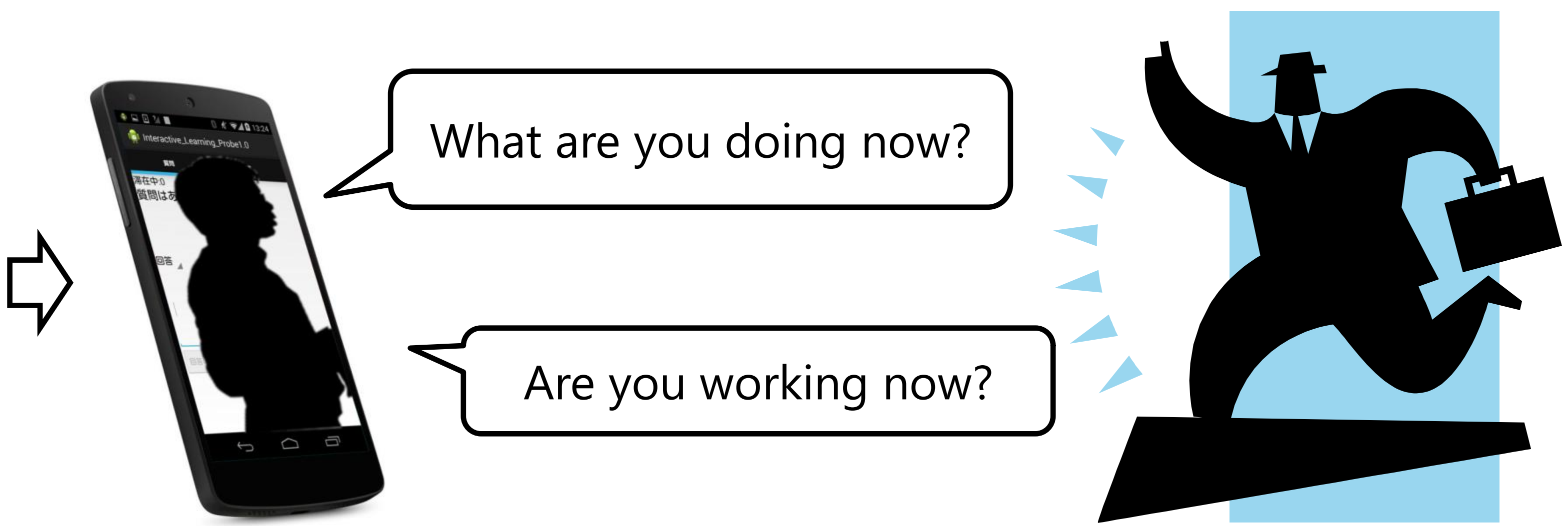
$$P(\hat{c}|Y)$$

Methodology

Concept

- The proposed method estimates activity type
- The system always measures activity situation using standard sensors
  - date, time, location
- The system detects occurrence of an activity
  - move-or-stay identification
- The system can ask a question about activity to the survey participants

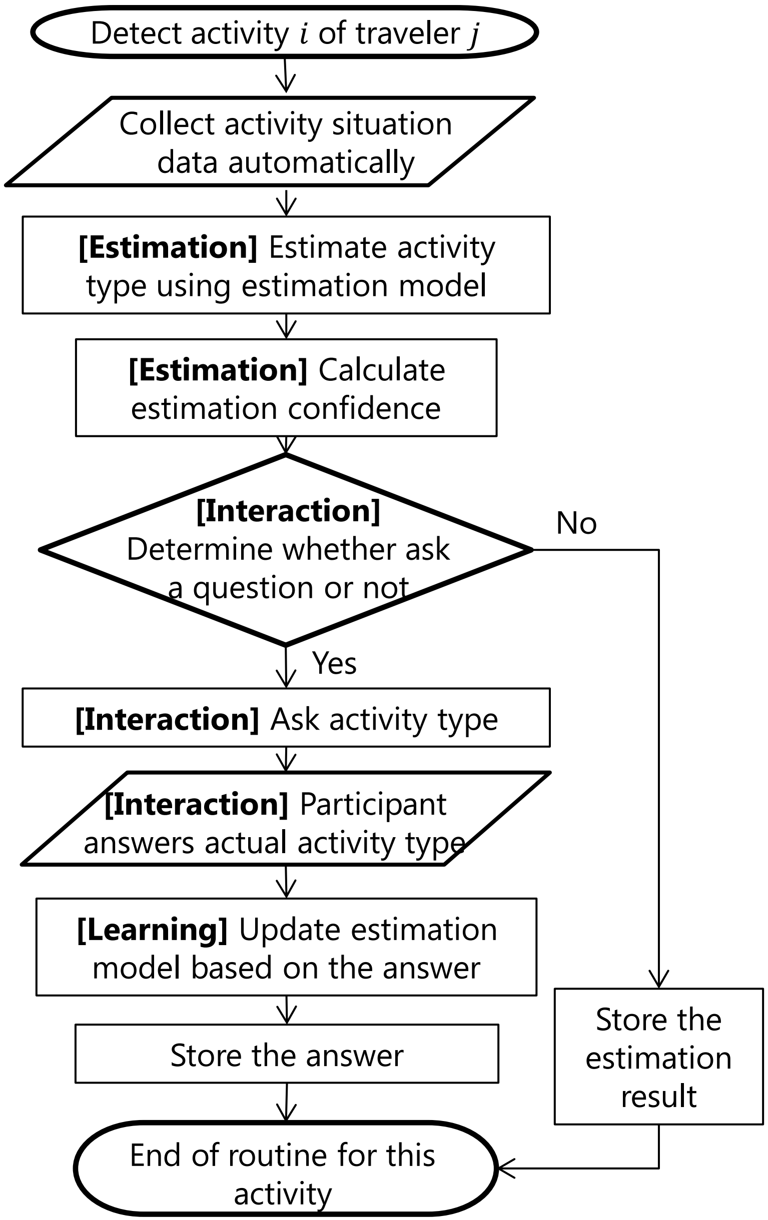
- date
- time
- location



Concept

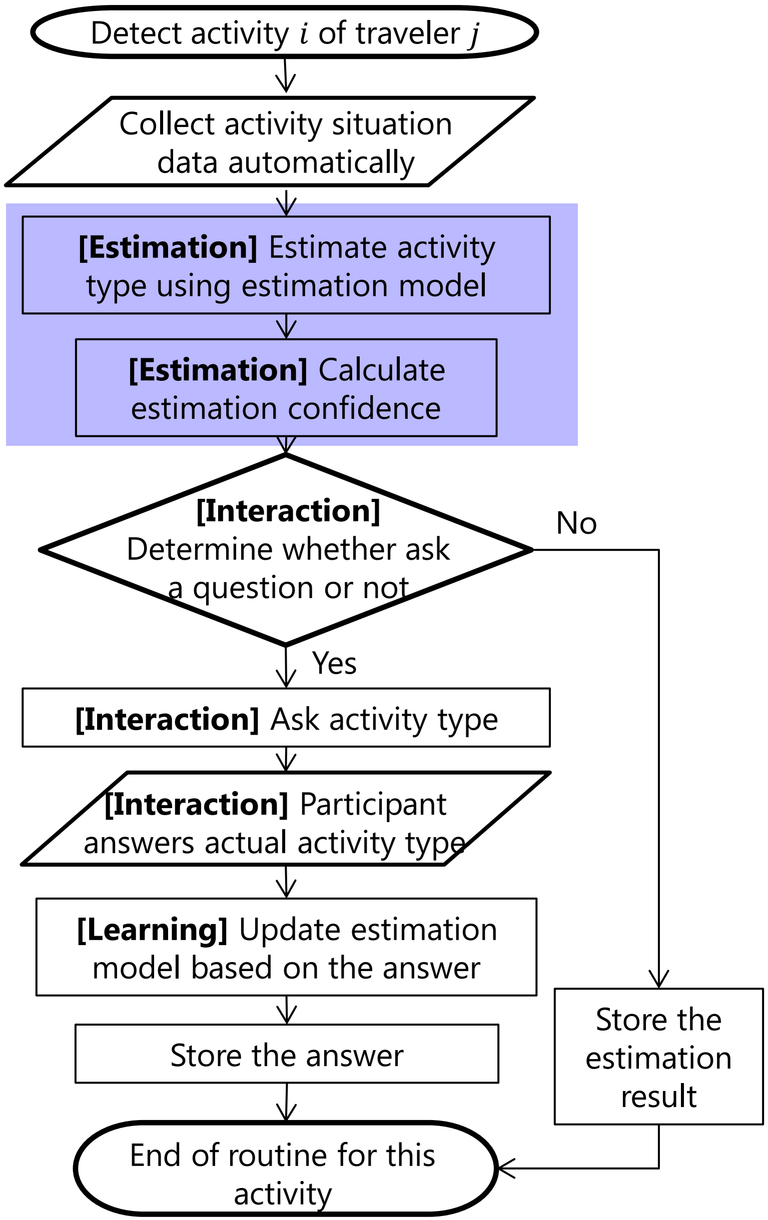
- The system can ask a question, if the system cannot estimate the activity
- The system learns a traveler-specific behavior pattern
- The system runs automatically

Overview



- Activity detection phase
  - Detects the survey participant staying somewhere to do unknown activity
  - out of scope of this study
- Estimation phase
  - Estimates the activity type using an traveler-specific estimation model
- Interaction phase
  - Will ask activity type to the survey participant, if the estimation is not confident
  - Will not ask the question, if the estimation is confident
- Learning phase
  - Updates the estimation model based on the participant's answer

Estimation phase



- Activity type estimation problem:

$$\hat{c} = \text{argmax}_c \, P(c|Y)$$

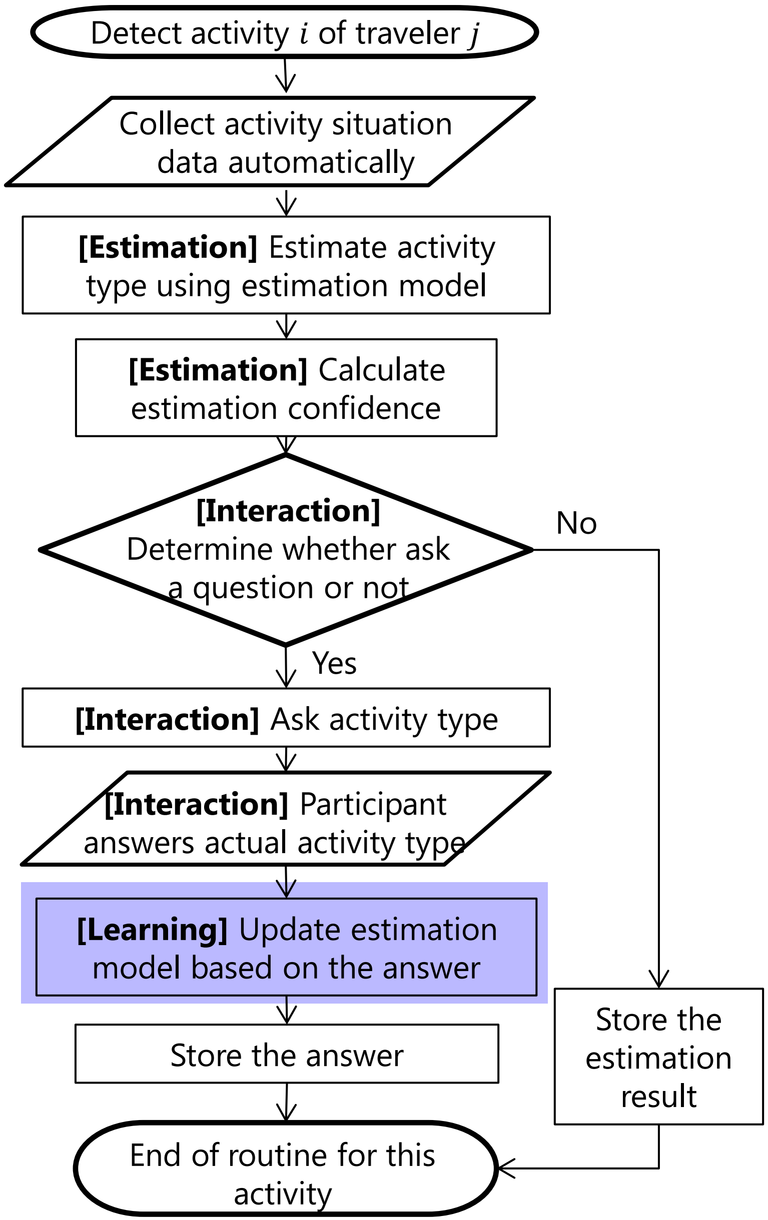
where
  - activity type: c
    eg: work, leisure
  - activity situation: $Y = (x_1, x_2, \dots)$
    eg: time: x_1, location: x_2

- Naive Bayes assumption:

$$\hat{c} = \text{argmax}_c \prod_k P(x_k|c)P(c)$$

- $P(x_k|c)$ and $P(c)$ are calculated based on historical data = learning

Learning phase



- Historical data:

$$H_{\text{traveler,time}} = \{(c_1, Y_1), (c_2, Y_2), \dots (c_i, Y_i), \dots\}$$

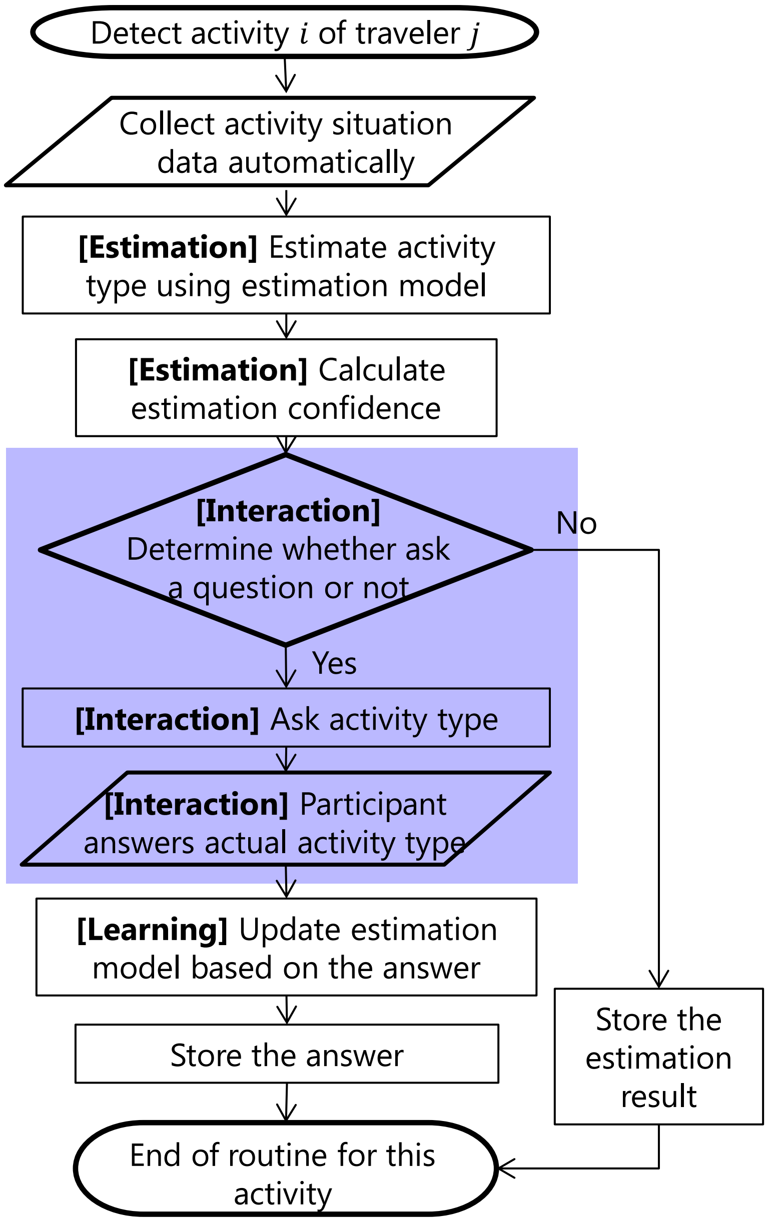
  - traveler-specific
  - dynamically updated

- $P(x_k|c)$ and $P(c)$ can be easily calculated based on the historical data

- How to collect the historical data?

  → online interaction between the system and participant

Interaction phase

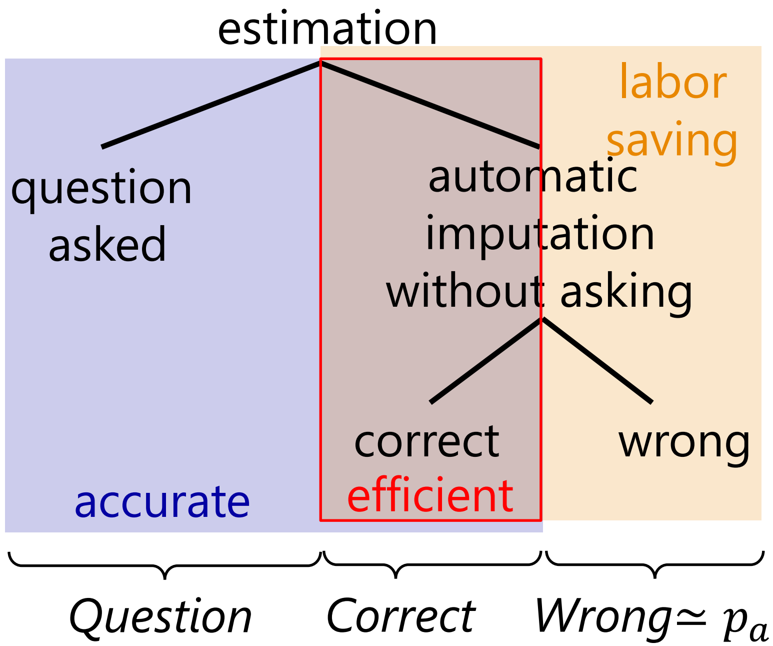


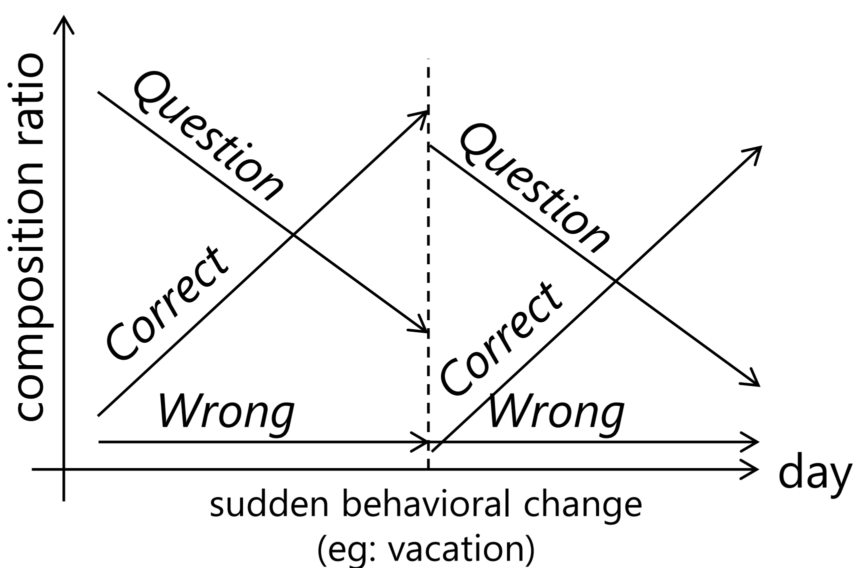
- Interaction:
  - If estimation confidence is high enough, the estimation result will be stored as a survey result
  - Otherwise, the system will ask the actual activity to the survey participant
    - The answer is stored to historical data for learning, as well as a survey result
- Estimation confidence:

$$P(\hat{c}|Y)$$

  - probability of performing activity $\hat{c}$ under situation Y
  - easily calculated based on historical data
- The system will ask question with certain probability p_q so that expected error rate will be equal to a given acceptable error rate p_a
  - p_a is given by the survey planner
    eg: 5%

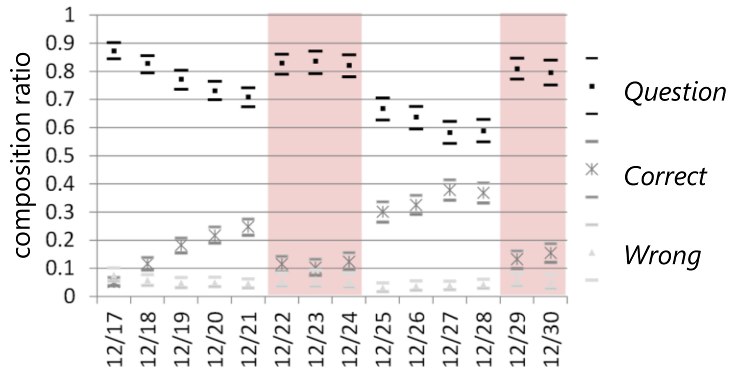
Summary





- The survey admin will be happy if *Wrong* is small
- The survey participants will be happy if *Question* is small
  - It makes long-term survey easier
- The acceptable error rate p_a is given by admin
  - quality control
  - trade-off: *Wrong*↓ ⇔ *Question*↑
- Estimation model keeps being updated during the entire survey period, for each travelers (=online)
  - At the initial stage, the estimation model is dumb
    - *Question* will be frequent
  - As the survey progresses, the estimation model will become accurate
    - *Question* will decrease; *Correct* will increase
  - Long-term behavioral change can be tracked
  - Traveler heterogeneity can be captured

2007/11/7	142	1000	0.666667	0.7916
2007/11/8	146	1000	0.641026	
2007/11/9	131	1000	0.625	0
2007/11/10	154	1000	0.818182	0.8260
2007/11/11	97	1000	0.75	0
2007/12/17	315	1000	0.785714	0.8928
2007/12/18	361	1000	0.833333	0.8571
2007/12/19	356	1000	0.818182	0.838
2007/12/20	369	1000	0.787879	0.8
2007/12/21	368	1000	0.75	0.718
2007/12/22	252	1000	0.837838	0.8333
2007/12/23	226	1000	0.851852	0.8148



Empirical Validation

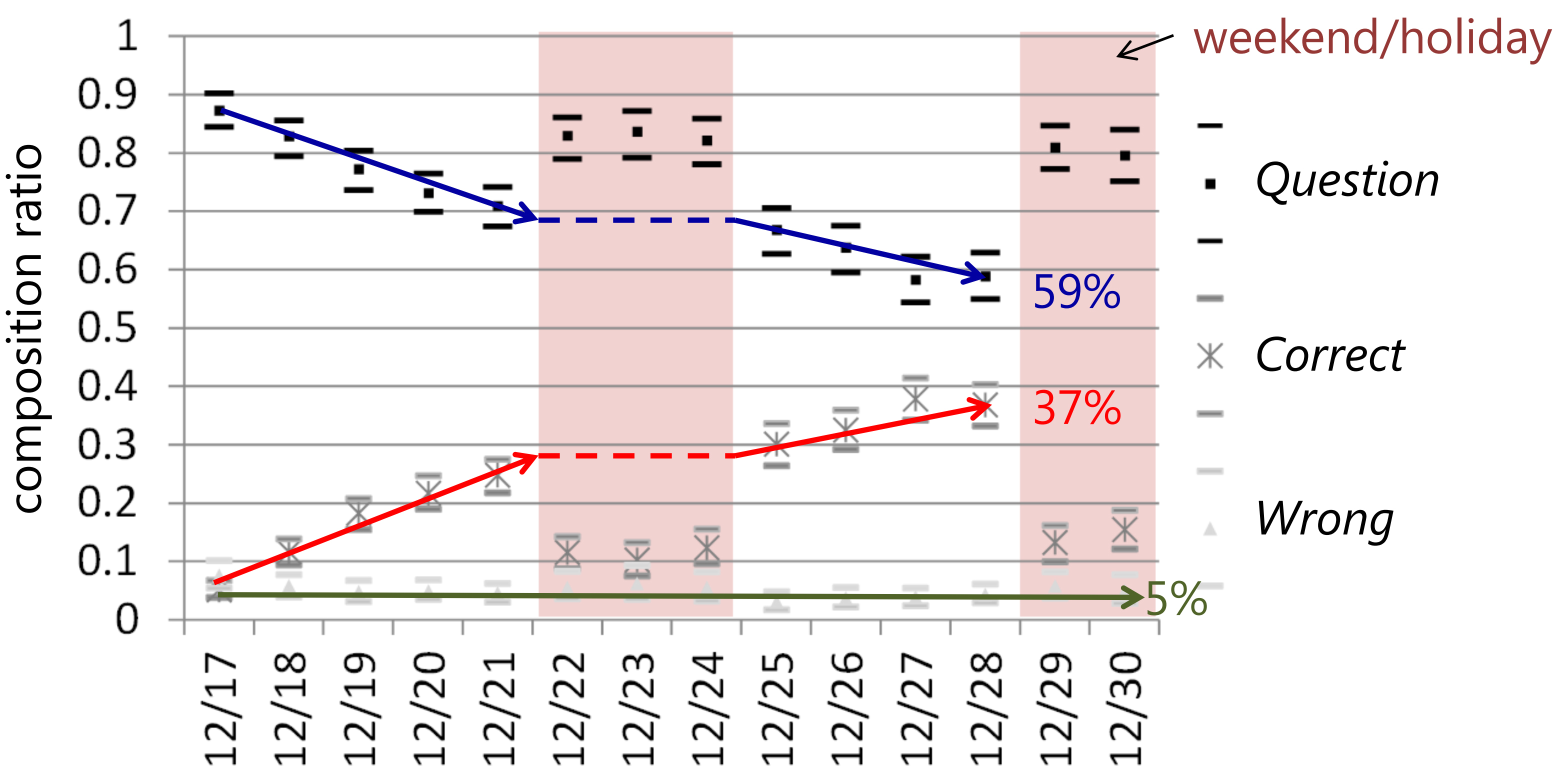
Data

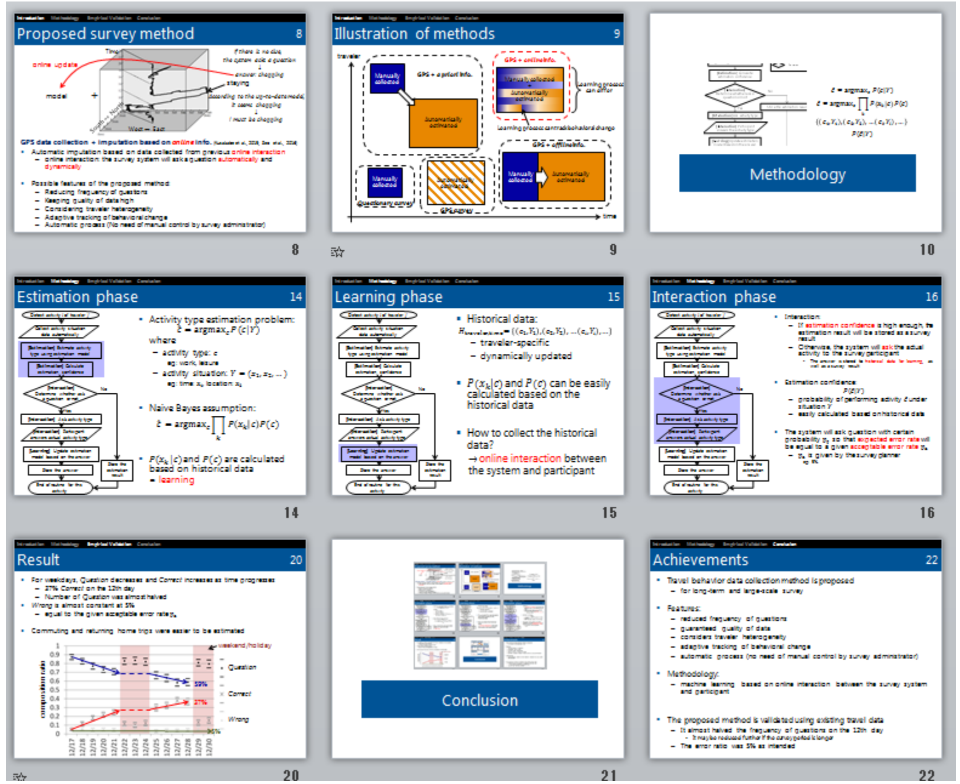
- The proposed method is validated by using existing travel survey data
  - collected by GPS + questionary survey
- Validation procedure
  1. Suppose that the survey data is true (ground truth)
  2. Emulate the proposed method
  3. Compare the estimation result with the ground truth

Date	Dec 17-30, 2007 • duration: 2 weeks
Location	Matsuyama city, Japan
Number of participants	92
Number of trips	4120 • 3.5 trips/person/day
Activity type (trip purpose)	commuting, returning home, business, shopping, food/leisure, others
Activity situation	weekday dummy, arrival time, location

Result

- For weekdays, *Question* decreases and *Correct* increases as time progresses
  - 37% *Correct* on the 12th day
  - Number of *Question* was almost halved
- *Wrong* is almost constant at 5%
  - equal to the given acceptable error rate p_a
- Commuting and returning home trips were easier to be estimated




Proposed survey method
8
Illustration of methods
9
Methodology
10
Estimation phase
14
Learning phase
15
Interaction phase
16
Result
20
Conclusion
21
Achievements
22

Conclusion

Achievements

- Travel behavior data collection method is proposed
  - for long-term and large-scale survey

- Features:
  - reduced frequency of questions
  - guaranteed quality of data
  - considers traveler heterogeneity
  - adaptive tracking of behavioral change
  - automatic process (no need of manual control by survey administrator)

- Methodology:
  - machine learning based on online interaction between the survey system and participant

- The proposed method is validated using existing travel data
  - It almost halved the frequency of questions on the 12th day
    - It may be reduced further if the survey period is longer
  - The error ratio was 5% as intended

References

Kitamura, R.: Panel analysis in transportation planning: An overview, *Transportation Research Part A: General*, Vol. 24, No. 6, pp. 401-415, 1990

Axhausen, K. W., Löchl, M., Schlich, R., Buhl, T. and Widmer, P.: Fatigue in long-duration travel diaries, *Transportation*, Vol. 34, No. 2, pp. 143-160, 2007

Asakura, Y. and Hato, E.: Tracking survey for individual travel behaviour using mobile communication instruments, *Transportation Research Part C: Emerging Technologies*, Vol. 12, No. 3-4, pp. 273-291, 2004

Hato, E.: Development of behavioral context addressable loggers in the shell for travel-activity analysis, *Transportation Research Part C: Emerging Technologies*, Vol. 18, No. 1, pp. 55-67, 2010

Cottrill, C. D., Pereira, F. C., Zhao, F., Dias, I. F., Lim, H. B., Ben-Akiva, M. E. and Zegras, P. C.: Future Mobility Survey, *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2354, No. 1, pp. 59-67, 2013

Wolf, J., Guensler, R. and Bachman, W.: 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*, No. 1768, pp. 125-134, 2001

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

Gong, L., Yamamoto, T. and Morikawa, T.: Comparison of Activity Type Identification from Mobile Phone GPS Data Using Various Machine Learning Methods, *Asian Transport Studies*, Vol. 4, No. 1, pp. 114-128, 2016

Kim, Y., Pereira, F. C., Zhao, F., Ghorpade, A., Zegras, P. C. and Ben-Akiva, M.: Activity Recognition for a Smartphone Based Travel Survey Based on Cross-User History Data, *2014 22nd International Conference on Pattern Recognition*, pp. 432-437, 2014

Kim, Y., Pereira, F. C., Zhao, F., Ghorpade, A., Zegras, P. C. and Ben-Akiva, M.: Activity recognition for a smartphone and web based travel survey, *arXiv preprint arXiv:1502.03634*, 2015

Kusakabe, T., Seo, T., Goto, H. and Asakura, Y.: Interactive Online Machine Learning Approach for Activity-Travel Survey, *Proceedings of the 14th International Conference on Travel Behaviour Research*, 2015

Seo, T., Kusakabe, T. and Asakura, Y: Formulation and Evaluation of Framework for Probe Person Survey using Online Machine Learning based on Interaction with Survey Participant, *Proceedings of Infrastructure Planning*, Vol. 54, 2016 (in Japanese)