

Collecting reliable training data to extract tourism strolling behaviour from smartphone GPS logs during walking

AI, Hisatoshi ^{a, *}, KAJI, Hideki ^a

^a Center for Spatial Information Science, The University of Tokyo, AI, hisaai@csis.u-tokyo.ac.jp, KAJI, kaji@csis.u-tokyo.ac.jp

* Corresponding author

Keywords: Tourism strolling behaviour, training data collection, smartphone GPS logs

Abstract:

A smartphone can be a useful device to deliver tourism information to users. Many earlier studies have discussed a method to determine whether the user is away from the area of their daily life, which would imply that the user is a potential tourist, or to select appropriate contents to be delivered to the user based on the user's preference and circumstances. However, few studies actually attempt to find the right timing for these tourism recommendations. The ultimate goal for this study is to develop a method to extract tourism strolling behaviour through real-time analysis of GPS log data collected from smartphones. We assume that the user will be inclined to visit a recommended spot if the information about the spot is received while strolling. This method will be useful to develop tourism spot recommendation applications or to equip current maps or navigation systems with the recommendation system.

We have developed a web application (Ai and Kaji, 2017; 2019) that watches location information collected from a smartphone GPS and analyses walking speed to provide a notification about nearby tourism spots only if the user is considered to be inclined to visit this spot. Figure 1 shows the interface of the web application. In an initial step, the user is asked to input a unique ID for data collection and to choose a target region. The web application starts to collect location information and analyse walking speed after the "Start logging" button is tapped. The data is collected once per second and is saved on the local storage of the browser. If the user's walking speed shows a certain pattern, the web application determines that the user is now strolling and shows tourism information on a screen. The user will evaluate the timing of the notification by tapping either a "Good" button or a "Bad" button. For this research, we added a "Request" button to express an interest in receiving a tourism recommendation when the web application does not show anything and neither the "Good" nor the "Bad" button is activated. These evaluations are also saved on local storage. If the user taps the "Send to server" button, data on the local storage will be sent to our server. By tapping the "End logging" button, the web application stops monitoring location information.

A proof-of-concept experiment is conducted in the field to collect training data, namely the evaluation taps, to improve the extracting algorithm of tourism strolling behaviour. We picked two target areas, Kawagoe and Yokohama; both cities are located in the Tokyo metropolitan area and have several tourism spots downtown. Participants are asked to walk from a railway station to a designated meeting place for two hours. During the walk, they were also requested to go through a shopping mall from the station and visit well-known tourism destination zones located between the shopping mall and the meeting place. Table 1 shows the date and participants' ID. The ID starts with a single letter, K or Y, meaning the participant only joined the experiment in Kawagoe or Yokohama respectively. The ID may include two letters, which means that the participant joined the experiment in both areas, and the order in the letters corresponds to that of participation, e.g. KY means first in Kawagoe and then in Yokohama.

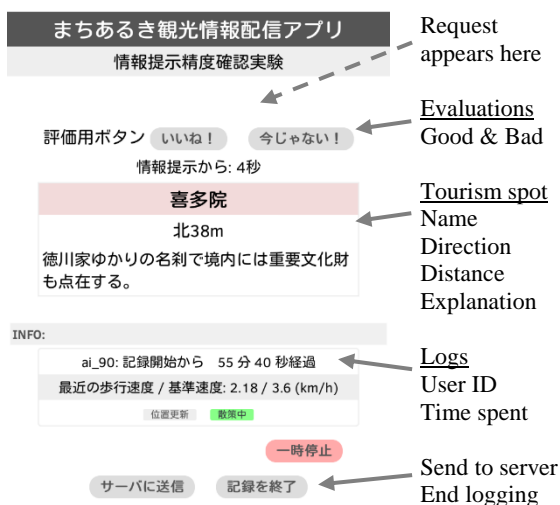


Figure 1. Interface of the web application

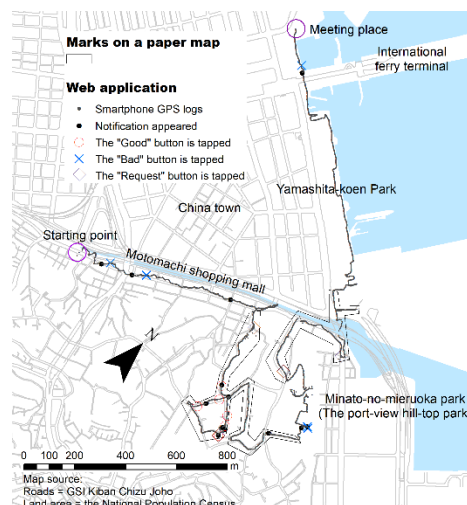


Figure 2. GPS log and training data collected by the experiment

Regions	Dates	Participants
Kawagoe	Sept. 27, 2018	K01 - K03, KY01 - 07
	Oct. 14, 2018	K04 - K06, KY08 - KY19
	Dec. 2, 2018	K07, YK01
Yokohama	Nov. 11, 2018	Y01 - Y04, KY01 - KY03, KY08 - KY14
	Dec. 1, 2018	Y05 - Y07, KY04 - KY07, KY15 - KY19, YK01

Table 1. Date and participants' ID of the proof-of-concept experiment

In this experiment, we collected two sets of training data. One comprises evaluation taps from participants, while the other contains hand-written marks on a paper map that show where participants thought they were strolling. Evaluation taps can be input in real time and on-site, however, “Good” or “Bad” evaluations can only be obtained while the web application identifies the participant as strolling. Although a participant can also tap the “Request” button, there is a limitation in collecting training data during the entire span of the experiment. On the other hand, marks on a paper map can cover the entire span of the experiment as it relies on a participant’s memory. However, there are potential limitations on the accuracy front. 1) Do participants perfectly remember where they were strolling for a whole two-hour walk? 2) Can participants correctly match the location they recalled strolling at to a location on the map? These two limitations can be overcome by using smartphone devices, since we can collect training data in real time and on-site. However, the use of a smartphone still introduces a limitation in data collection intervals as mentioned before.

In this presentation, we will discuss potential differences between the two training data sets – namely evaluation taps on a smartphone and marks on a paper map.

Table 2 shows the summary for training data. Rows marked G, B, and R show how many taps are made on “Good,” “Bad,” and “Request” buttons respectively. “Match” rows to the right of G, B, and R show how many taps do not conflict with marks on a paper map, e.g. number of “Good,” or “Request” taps within the marks, and number of “Bad” outside the marks. The row labelled “Rate” shows the overall rate of taps that matched marks on a paper map, e.g. total match counts divided by total number of taps. Data is missing on K05 following an error transferring data to our server. Only few participants achieved a high rate even when they made many taps, e.g. Y07 and KY19. We also found that for some participants, such as KY03 and KY10, the rate varies in the two areas even when they originated from the same person. This suggests that the evaluations are vague and that the methodology used to efficiently collect reliable training data is key as we try to improve the timing of recommendations.

	Kawagoe								Yokohama							
	ID	G	Match	B	Match	R	Match	Rate	ID	G	Match	B	Match	R	Match	Rate
Participated only in Kawagoe or in Yokohama	K01	2	0	3	3	1	0	50	Y01	3	0	8	8	3	0	57.1
	K02	1	0	0	0	4	1	20	Y02	9	1	13	11	7	2	48.3
	K03	3	1	11	3	2	1	31.3	Y03	5	3	8	2	1	1	42.9
	K04	17	14	8	0	2	2	59.3	Y04	11	9	7	1	8	5	57.7
	K05								Y05	9	2	4	2	0	0	30.8
	K06	8	4	21	16	8	0	54.1	Y06	22	8	18	12	1	1	51.2
	K07	3	0	0	0	6	4	44.4	Y07	17	15	2	0	18	16	83.8
Participated in both regions	KY01	15	9	8	8	15	8	65.8	KY01	7	4	9	5	4	2	55
	KY02	9	3	2	1	1	1	41.7	KY02	8	2	4	2	1	0	30.8
	KY03	3	2	12	10	0	0	80	KY03	6	1	18	9	0	0	41.7
	KY04	0	0	1	0	1	0	0	KY04	0	0	2	1	1	0	33.3
	KY05	5	0	6	6	1	0	50	KY05	7	4	8	7	0	0	73.3
	KY06	3	2	0	0	1	1	75	KY06	4	4	0	0	0	0	100
	KY07	4	0	10	6	2	0	37.5	KY07	5	3	12	0	0	0	17.6
	KY08	5	0	0	0	3	1	12.5	KY08	1	1	1	0	5	0	14.3
	KY09	5	3	10	8	5	0	55	KY09	9	7	26	1	0	0	22.9
	KY10	2	2	0	0	13	9	73.3	KY10	5	0	14	9	0	0	47.4
	KY11	0	0	1	1	0	0	100	KY11	0	0	6	3	0	0	50
	KY12	6	6	10	2	5	2	47.6	KY12	2	1	1	1	4	3	71.4
	KY13	3	3	8	0	1	1	33.3	KY13	2	1	0	0	11	7	61.5
	KY14	9	5	6	3	4	1	47.4	KY14	9	7	13	1	11	10	54.5
	KY15	4	1	1	1	0	0	40	KY15	0	0	0	0	0	0	-
	KY16	15	7	10	3	4	1	37.9	KY16	11	4	4	1	8	4	39.1
	KY17	8	2	9	8	2	1	57.9	KY17	6	0	8	7	1	0	46.7
	KY18	10	2	16	12	1	0	51.9	KY18	2	0	9	8	0	0	72.7
	KY19	21	16	14	6	1	1	63.9	KY19	6	6	5	5	4	4	100
YK01	2	1	0	0	1	1	66.7	YK01	1	0	0	0	4	0	0	

Table 2. Number of taps on evaluation buttons and how many of them are made while inside the marks on a paper map

Reference:

AI, Hisatoshi, and KAJI, Hideki (2017) Improving a web application that delivers tourism information to the user at good timing, *Research Abstracts on Spatial Information Science CSIS DAYS 2017*, B11. In Japanese.

AI, Hisatoshi, and KAJI, Hideki (forthcoming) Basic analysis of timing of tourism information notification to walking users -a demonstration experiment using web application-, *Society for Tourism Informatics*, 15(1). In Japanese.