Towards Gateless Railway Services using GPS Location Based Ride Detection

Jun Nemoto

Graduate School of Science and Technology
Keio University
Yokohama, Japan
Email: nemoto@keio.jp

Motomichi Toyama

Faculty of Science and Technology
Keio University
Yokohama, Japan
Email: toyama@ics.keio.ac.jp

Abstract—**Gateless ticket inspection is an interesting and attractive feature that can help achieve less waiting, less fare evasion, less difficulty of use for railway services for everyone, including physically handicapped people. In this paper, we study ride route detection using Global Positioning System (GPS) as a new approach for implementing gateless railway services. We assume the gateless railway service has access to the user's GPS location through an application on their smartphone or another mobile device. This position can then be compared with the GPS location of trains in order to detect the stations at which the user boarded and disembarked. Then, railway operators can charge the user for the ride. A challenge in the ride detection for fare charge in railway services is to detect the ride correctly, even if the GPS trajectory is short, e.g., in case users only ride for one station. In order to solve this challenge, we propose a ride detection solution that uses different criteria to evaluate how far the user and train were moving between two stations. In our simulation using railway line open data provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) in Japan, we show that our proposed method reduced false positives by 25%-100% in most cases.**

Keywords–GPS; GIS; Railway Ride Detection.

I. INTRODUCTION

Gateless ticket inspection is an interesting and attractive feature that can help achieve less waiting, less fare evasion, less difficulty of use of railway services for everyone, including physically handicapped people. For example, a barrier-free fare collection system using wireless communication technology has been proposed in [1]. It enables wheelchair users and people with strollers or big luggage to pass smoothly without using a smart card or ticket at the gate.

In this paper, we study ride detection using GPS location as a new approach for implementing gateless railway services. We assume the gateless railway service has access to the user's GPS location through an application on their smartphone, or another mobile device. This position can then be compared with the GPS location of trains in order to detect the stations at which the user boarded and disembarked. Then, railway operators can charge the user for the ride.

There has been a lot of research on transportation mode detection using GPS location [2]–[7]. These works have proposed to infer a user's mode of transportation, such as walking, car, and rail based on the velocity calculated by trajectories of GPS location, data of accelerometers, Geographic Information System (GIS) data, and so on. However, all of them do not discuss ride detection but focus on transportation mode detection. In the ride detection, we need to know not only that a user rode a train but exactly which train the user rode because, for example, two operators could run trains on the same line.

A challenge in ride detection for fare charging in railway services is to detect the ride correctly even if the GPS trajectory is short. It is not difficult to infer the transportation mode if a user's GPS trajectory is relatively long across multiple stations, as noise in the data can be averaged out, and data preceding and following a time point can be used for inference (e.g., if a user rode a train between A and B and C and D they probably also rode it between B and C). A gateless railway service needs to accurately charge all rides, even short ones so we take the harder case of short GPS trajectories into particular consideration.

To address the challenge of accurate ride detection, we propose a ride detection system based on the GPS position of both the user and the train. Using the estimated distance between the user and the train, our system is able to accurately detect rides, even when they occur only between two stations.

The rest of this paper is organized as follows. First, we provide related works in Section II. Then, we introduce the overview of the gateless railway service in Section III. We detail the proposed method in Section IV and evaluates it by the simulation in Section V. In Section VI, we discuss the challenges to extend our approach to the real world. Finally, we summarize the conclusions in Section VII.

II. RELATED WORK

Many studies in the previous decade have focused on inferring transportation modes based on GPS location. Some of them tried to infer the modes only from GPS trajectory data [2][7] while others tried to improve the accuracy of the inference by using additional information, such as accelerometer data [5] and GIS data [3][4][6]. However, all of them are not discussing the ride detection but are focusing on transportation mode detection.

As we mentioned in Section I, the main difference between the transportation mode detection and the ride detection is that high accuracy is required even for short GPS trajectories. When considering ride detection, it is necessary for fare charging to infer whether a user rode the train or not, even if they only rode it just for one station. However, previous studies and the criteria and features used in them tend not to account for such short GPS trajectories. Thus, applying them does not provide good enough accuracy of the inference in the ride detection.

For example, Stenneth et al. [3] proposed a machine learning approach using 8 features to infer the transportation

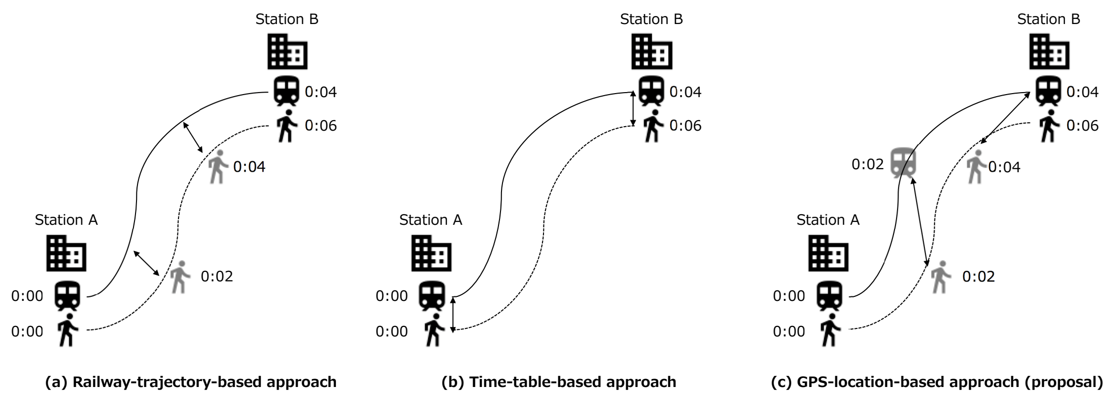

Figure 1. Comparison of railway ride detection

mode detection. In these features, the average distance between a user and a railway trajectory can be applied to the railway ride detection. Hereafter, we call the ride detection using this feature *railway-trajectory-based approach* in this paper. Figure 1(a) shows a conceptual diagram of the railway-trajectory-based approach. Since this approach does not consider the train's location but only uses the distance between a user and the closest railway line, decision errors might occur frequently if a road runs along the railway line, as the figure shows.

Montoya et al. [6] proposed an approach that considers the train location. They use station locations and timetables to distinguish train, subway, and tram in the transportation mode detection. More specifically, using route information in General Transit Feed Specification (GTFS) format provided by railway operators, they use the average distance between a user and stations on the route. In addition, they consider departure time and arrival time based on the timetable of the route. Hereafter, we call the ride detection using these criteria *timetable-based approach* in this paper. Figure 1(b) shows a conceptual diagram of the timetable-based approach. However, this approach has a limitation in the accuracy of the ride detection for short GPS trajectories since it does not consider the train location between stations.

To solve these challenges, we propose a *GPS-location-based approach* that uses GPS locations of both users and trains for the ride detection. Our proposed approach is based on the distance between the user and the train between two stations, based on their GPS locations. Figure 1(c) shows a conceptual diagram of GPS-location-based approach and its details are described in Section IV.

Note that Stenneth et al. [3] also use 2 features based on real-time bus locations but both of them are not appropriate for ride detection. One of them is Average Bus Closeness (ABC) that is the average distance between a user and the closest bus in a given time series, and the other one is Candidate Bus Closeness (CBC) that is the minimum value among the sums of the distance between a user and a bus at each time in a given time series. The former is a feature that mixes information regarding multiple buses and the latter is a feature that mixes information across the multiple bus stops. Thus, both of them cannot be used for ride detection.

III. Gateless Railway Service

This section describes the gateless railway service that our system could be used in. Figure 2 shows the overall architecture of the service.

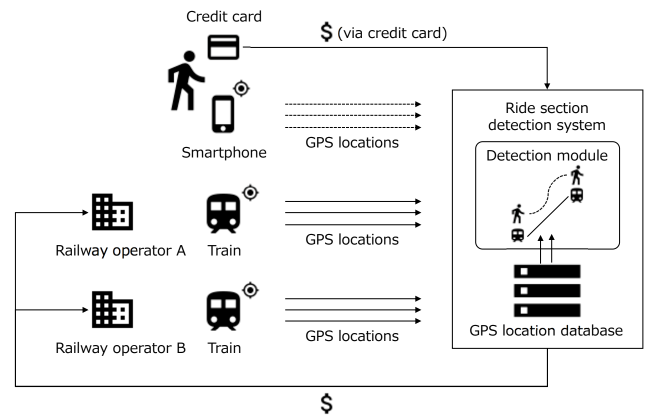

Figure 2. Gateless railway service architecture

In this gateless railway service, the ride detection system regularly collects GPS locations from user's devices, such as smartphones, and from trains operated by railway operators, and stores them in a database. The ride detection system periodically creates ride histories for each user and charges a user's credit card according to the histories. Then, the collected payments will be paid to the railway operators.

Introducing the gateless railway service which removes the gate itself has several advantages for both users and railway operators though smooth ticket examining using contactless smart cards have already been achieved in many countries. For users, the accessibility in stations will be drastically improved. For example, wheelchair users and people with strollers or big luggage can access platforms smoothly without minding the narrow gate. In addition, although a long queue can be formed at the time of congestion in main stations, the gateless railway service can alleviate it. As for railway operators, reducing costs for introducing and maintaining automatic ticket examining machines can be expected. Also, fare evasion in unmanned stations can be reduced if a fare control based on the degree of contribution for recording GPS locations as described in Section VI and sudden ticket examination is combined and performed.

IV. Railway Ride Detection

This section describes the proposed ride detection method.

Let l_t^p and l_t^q be the location of a train p and a user q at a time t, respectively. We say that the user q is in *presumed* *ride state* for the train p if Euclidean distance d between l_t^p and l_t^q is less than or equal to the threshold θ_d.

Let $L^p_{(A,B)}$ be a sequence of the train p's location between stations A and B.

$$L^p_{(A,B)} = \{l_1^p, l_2^p, \cdots, l_n^p\} \tag{1}$$

where l_1^p is the train p's location at the departure time and l_n^p is the train p's location at the arrival time.

Similarly, a sequence of the user q's location at the same time can be represented as follows.

$$L^q = \{l_1^q, l_2^q, \cdots, l_n^q\} \tag{2}$$

We call the proportion of being the presumed ride state in n judgments *presumed ride rate* for the user q in $L^p_{(A,B)}$. We assume that the railway operators determine that the user virtually rode in the section and charge when the presumed ride rate is greater than or equal to a certain threshold.

V. Evaluation

In this section, we evaluate our GPS location-based approach with respect to the accuracy of the ride detection by comparing it with the railway-trajectory-based one and the timetable-based one in a simulation. First, we describe the methodologies of the evaluation in Section V-A and then show the results in Section V-B.

A. Methodologies

1) Indicators: There are the following 4 patterns for the results of the inference.

- True Positive:
  Inferred virtually rode and actually rode.
- True Negative:
  Inferred not rode but actually not rode.
- False Positive:
  Inferred virtually rode but actually not rode.
- False Negative:
  Inferred not rode but actually rode.

We consider two indicators to evaluate the accuracy of the ride detection based on the number of each case above.

$$r_{FP} = \frac{N_{FP}}{N_{TN} + N_{FP}} \tag{3}$$

$$r_{FN} = \frac{N_{FN}}{N_{TP} + N_{FN}} \tag{4}$$

where r_{FP} is the false positive rate, r_{FN} is the false negative rate, and N_{TP}, N_{TF}, N_{FP}, N_{FN} are the numbers of true positives, true negatives, false positives, and false negatives respectively.

The false positive rate can be used to evaluate the possibility that users need to pay fare unreasonably and the false negative rate can be used to evaluate the possibility that railway operators will fail to collect the estimated fare. Of the two indicators, the false negative rate can be adjusted by changing the distance threshold θ_d, which is used for judging the presumed ride state. Thus, in this simulation, we evaluate the false positive rate by using the cases that can be misjudged as virtually rode.

TABLE I. RAILWAY OPERATORS
AND EXAMPLES OF RAILWAY LINES

Railway operators	*Number of lines in the evaluation*	*Examples*
Odakyu	1	Odawara Line
Keio	7	Keio Line, Inokashira Line
Keikyu	1	Main Line
Seibu	2	Shinjuku Line, Ikebukuro Line
Tokyu	2	Toyoko Line, Denentoshi Line
TWR	1	Rinkai Line
Tobu	2	Isesaki Line, Tojo Line
JR East	34	Yamanote Line, Tokaido Line

2) Train Location: For the simulation, we use pseudo location information based on the railway trajectory data and the timetable data instead of the actual GPS location. MLIT in Japan provides GIS data, such as railway trajectories. Using the railway trajectories provided by MILT and the duration in timetables, we calculate the location of the train at a certain time. Specifically, assuming the train moves with a constant speed, we evenly divide the trajectory curve between stations by the distance moved at a fixed time interval (10 seconds in this evaluation).

3) User Location: We assume that a user travels between the target stations by car and unintentionally causes the misjudgment because the roadway often runs parallel nearby the railway line in urban areas of Japan. Specifically, we use the trajectory curve of the recommended drive route and its duration obtained by Google Maps Application Programming Interface (API). When calculating the location of the user, we assume that the car runs at a constant speed, for the simplicity.

4) Railway Lines: The targets are 50 lines and total 858 sections in the suburbs of Tokyo, Japan, which are operated by 8 railway operators. These 8 operators and example railway lines are shown in Table I. Note that we use one of the lines for the evaluation if an operator runs multiple lines in the same section. In addition, we exclude subway due to the difficulty of obtaining GPS locations but its detail will be discussed in Section VI.

B. Results

First, we compare the false positive rate between the railway-trajectory-based approach and our proposal while varying the threshold of the presumed ride rate and the distance (θ_d) in order to evaluate how the rate will be improved when considering the train position.

As shown in Table II, the proposed method provides a highly accurate false positive rate (about 1%) when lowering the presumed ride rate to 0.8 and raising the threshold of the distance to 150m.

On the other hand, in the railway-trajectory-based approach, misjudgments occurs in 5% of the sections even if the threshold of the distance is set to 50m. The false positive rate of the railway-trajectory-based approach may increase further in the real world situation since the length of the train is about 150-300m in general and railway operators would like to raise the threshold of the distance based on it in order to improve the false negative rate.

Next, we compare the false positive rate between the timetable-based approach and our proposal in order to evaluate how the rate will be improved when judging the position of

 ISBN: 978-1-61208-700-9

TABLE II. COMPARISON OF FALSE POSITIVE RATE WHETHER TRAIN POSITION IS CONSIDERED OR NOT

Threshold of presumed ride rate	*Threshold of distance (m)*	*Railway-trajectory-based*	*Proposal*
0.8	50	5.4%	0.0%
	100	17.0%	0.3%
	150	30.3%	1.0%
0.9	50	2.8%	0.0%
	100	10.8%	0.2%
	150	21.8%	0.8%
1.0	50	0.6%	0.0%
	100	5.2%	0.2%
	150	14.5%	0.6%

TABLE III. COMPARISON OF FALSE POSITIVE RATE WITH CONSIDERING TRAIN DELAY (THRESHOLD OF PRESUMED RIDE RATE ≥ 0.8)

Delay (sec)	*Threshold of distance (m)*	*Timetable-based*	*Proposal*
30	50	0.9%	0.1%
	100		0.6%
	150		2.4%
60	50	2.4%	0.0%
	100		1.2%
	150		4.1%
120	50	7.5%	0.0%
	100		1.9%
	150		5.6%

users and trains between stations in a fine-grained manner. Table III shows the results.

In this evaluation, since we assume the distance between the user and the departure/arrival station is zero, the timetable-based approach practically infers based on only the difference of the time duration between the user and the train. In other words, the false positive rate will vary only according to the tolerance for the time difference. For example, when using the low threshold of the time difference, the false positive rate will be improved but many false negative cases occur in the case of train delay. Thus, we vary the threshold of the time difference with 30, 60 and 120 seconds and we show the false positive rate of each case in Table III. Note that the rate in the case of the train delayed by 30, 60 and 120 seconds is shown for the proposed approach.

As shown in Table III, the proposed method can achieve good accuracy as a whole though it falls into a higher false positive rate than the timetable-based approach when using the large threshold of the distance.

VI. DISCUSSION

In this paper, we propose a ride detection method based on the GPS locations in the gateless railway service. However, there are many challenges to be overcome in the production system.

A. Radio Wave Dead Zones

Applying our GPS-location-based approach to radio wave dead zones, such as subways and tunnels is a challenge. However, nowadays, mobile devices can receive radio waves even if they are in subways or tunnels and can infer the location according to the cellar base stations and WiFi access points. The accuracy is not high (e.g., a few kilometers), but we believe that it is possible to detect the ride section based on not only the GPS locations but also the data from other sensors, such as an accelerometer.

B. Countermeasures for Fare Evasion

Investigating the countermeasures for the fare evasion is also a challenge. As a simple way of the fare evasion, just turning off the power of the smartphone can be considered. Using faked GPS locations is also possible in some way. The former type of cheat can be reduced by introducing a fare control based on the degree of contribution for recording GPS locations. For example, offering some incentives by applying a higher fare if the GPS locations are intermittently lost may reduce the fare evasion. The latter type of cheat can be excluded by checking whether the operating system and/or the application is faked or not using API, such as SafetyNet in Android. However, the fare evasion cannot be detected if always turning off the power from the begining. Therefore, manual approaches, such as random control by service personnel and the expensive fine will be necessary.

VII. CONCLUSION

In this paper, we study ride detection based on GPS location as a new approach for implementing gateless railway service. Unlike the transportation mode detection, it is necessary for fare charging to detect the ride correctly even if the GPS trajectory is short. In order to achieve this, we proposed a ride detection method that uses criteria to evaluate how far the user and train were moving between two stations. In our simulation using railway line open data provided by MLIT in Japan, we show that our proposed method reduced false positives by 25%-100% in most cases.

We plan to develop a prototype and evaluate the proposed method with overall criteria including false negative error rate and efficacy of countermeasure for fraud prevention, such as incentive management by fare control.

ACKNOWLEDGMENT

The authors would like to thank Thomas Laurent for carefully proofreading the manuscript.

REFERENCES

[1] Mitsubishi Electric Corporation, "Concepts for Advanced Rail Travel," 2017, URL: http://www.mitsubishielectric.com/news/2017/pdf/1120-a.pdf [accessed: 2019-02-01].

[2] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Ma, "Understanding mobility based on gps data," in Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp 2008), 2008, pp. 312–321.

[3] L. Stenneth, O. Wolfson, P. S. Yu, and B. Xu, "Transportation mode detection using mobile phones and gis information," in Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2011, pp. 54–63.

[4] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, "A GPS/GIS method for travel mode detection in New York City," Computers, Environment and Urban Systems, vol. 36, no. 2, 2012, pp. 131–139.

[5] P. Widhalm, P. Nitsche, and N. Brndie, "Transport mode detection with realistic smartphone sensor data," in Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012, pp. 573–576.

[6] D. Montoya, S. Abiteboul, and P. Senellart, "Hup-me: Inferring and reconciling a timeline of user activity from rich smartphone data," in Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2015, pp. 62:1–62:4.

[7] Z. Xiao, Y. Wang, K. Fu, and F. Wu, "Identifying different transportation modes from trajectory data using tree-based ensemble classifiers," ISPRS International Journal of Geo-Information, vol. 6, no. 3, 2017, p. 57.