Trajectory Data Acquisition via Private Car Positioning Based on Tightly-coupled GPS/OBD Integration in Urban Environments

Zhu Xiao, *Senior Member, IEEE*, Yanxun Chen, Mamoun Alazab, *Senior Member, IEEE*, and Hongyang Chen, *Senior Member, IEEE*

***Abstract*— The explosive growth of road vehicles especially the private cars has brought unprecedented pressure to a series of problems in urban transportation systems, such as traffic congestion and environmental pollution. Private cars trajectory data and perceiving their information provide a promising solution to these problems. However, the collection of large-scale trajectory data for private cars with high accuracy and reliability is still delicate tasks in urban environments. In this paper, we propose a low-cost and user-friendly implementation method for achieving large-scale private cars trajectory data acquisition via designing lightweight GPS module and On Board Diagnostics (OBD) reader. To ensure reliable trajectory data acquisition via GPS/OBD integration, we propose an ensemble learning based Gauss Process Regression (GPR) method so as to cope with the non-linearity, non-stationarity and incremental training problems during trajectory collection. We design a classification-type loss (CTL) function and build a regression to classification (R2C) method with Learn++ for realizing ensemble learning. The proposed approach implements incremental learning when new trajectory data arrives and is able to resolve the concept drifting problem. Experiments in real-world urban environment have demonstrated the effectiveness and reliability of the proposed method, it achieves better trajectory prediction performance than the comparative methods under various road conditions in GPS-denied areas.**

***Index Terms*— GPS/OBD integration, private cars, trajectory data acquisition, vehicle positioning.**

I. Introduction

NOWADAYS, the ever-increasing of road vehicles pose great challenges on transportation management system and cause economic and environmental problems in modern cities [1]–[3]. In particular, the private cars, which are referred as small motor vehicles usually registered by individual and for personal use, consist of the vast majority of urban vehicles in the transportation systems. Taking China as an example, the nationwide ownership of civilian automobiles has reached 250 million by 2018 [4], 82.80% of which are registered by individual, namely, private cars reach 207 million. Therefore, large number of private cars traveling in the city produce huge volume of trajectory data. Driven by the massive number of private cars, their trajectory data can be utilized in the intelligent transportation systems (ITS) for the analysis of driving behaviors and transportation conditions [5], [6], which fosters various data-intensive applications ranging from vehicle infotainment to social big data analytics [7]–[11].

Existing studies have explored the trajectory data information collected from road vehicles in many applications. The majority of these studies focus on investigating the trajectory of the floating cars [12], namely, trajectory data created by buses and taxis [13]–[15]. However, the industry is facing the challenge, i.e., how to obtain vehicle trajectory data from the daily driving especially for large-scale private cars. This is mainly because the trajectory data of the floating cars are easily available while the collection of trajectory data from private cars is relatively scarce. Due to the need of management and the concern of road safety, taxi and bus are mandatory to install tracking devices such as GPS, so that the taxi company and municipal traffic authorities can monitor their driving trajectory. Nevertheless, this compulsive regulation is inapplicable to private cars. Therefore, how to realize trajectory data acquisition for private cars and ensure the accuracy and reliability of trajectory collection are still challenging tasks especially in complex urban environments. For example, the communication signals of different frequency bands and rates from various communication devices in cities produce interference to the GPS signals; Multi-story buildings and other obstacles in urban areas may cause blockage to GPS signal reception. These inevitably lead to the complex and GPS-denied urban environments for the trajectory data acquisition.

A. Related Works

With the proliferation of various smart devices such as smartphone wherein GPS and inertial sensors are embedded [16], it is technically likely to implement trajectory collection if the drivers use their own phone's positioning and navigation function during driving. However, most drivers are quite

Manuscript received 15 May 2021; revised 9 July 2021; accepted 13 August 2021. Date of publication 30 August 2021; date of current version 8 July 2022. This work was supported in part by the Humanities and Social Sciences Foundation of MOE under Grant 21YJCZH183, in part by the National Natural Science Foundation of China under Grant U20A20181 and Grant 61702175, in part by the Key Research and Development Project of Hunan Province of China under Grant 2022GK2020, in part by the Funding Projects of Zhejiang Lab under Grant 2021LC0AB05 and Grant 2020LC0PI01, and in part by the Science and Technology Project of Hunan Provincial Water Resources Department under Grant XSKJ2021000-39. The Associate Editor for this article was S. Mumtaz. *(Corresponding author: Hongyang Chen.)*

Zhu Xiao and Yanxun Chen are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: zhxiao@hnu.edu.cn; chenyanxun@hnu.edu.cn).

Mamoun Alazab is with the College of Engineering, IT and Environment, Charles Darwin University, Darwin, NT 0810, Australia (e-mail: mamoun.alazab@cdu.edu.au).

Hongyang Chen is with Zhejiang Lab, Hangzhou 311121, China (e-mail: dr.h.chen@ieee.org).

Digital Object Identifier 10.1109/TITS.2021.3105550

aware of the driving directions in daily city driving and they don't need the assistant from cellphone navigation unless they occasionally drive to an unknown location. If drivers use the smartphone navigation App to collect trajectory data, they need to click the App once they start to drive the car, since most people would not tend to keep the navigation App online. As such, people should avoid distraction from using phone, this in turn restricts the feasibility of trajectory collection. For these reasons, using smartphone is not a practical way for private cars to achieve long-playing trajectory data collection.

Vehicle positioning methods with help of vision-based technology [17], [18], inter-vehicle communication and RFID technology [19]–[22], can implement trajectory collection and work properly under urban areas. In [23], the authors investigate the vehicle trajectory prediction by combining physics & maneuver-based approach with IMM model. Based on the method of sequence analysis, the authors in [24] propose the kernel variable length Markov model (KVLMM) to address the trajectory prediction problem. Nevertheless, these methods are not applicable for large-scale trajectory data acquisition when considering the cost and operability. For instance, deploying wireless inter-vehicle connection for road vehicles is impractical at the current stage. The vision-based positioning such as LiDAR is attractive for developing autonomous intelligent vehicles but cost-prohibitive and hence hard to popularize to private cars.

A viable and practical solution aiming at realizing trajectory data acquisition, is to leverage the integration of external sensors and measurements, such as GPS and inertial measurement unit (IMU) devices [25], [26]. In this context, the integration of GPS and inertial navigation system (INS) has been applied widely in vehicle positioning and state prediction [27], [28]. With the data fusion concept, the authors in [29] utilize the rotation modulation, kinematic constraint (KC), and the robust adaptive Kalman filter to design a vehicle positioning scheme, which can reduce the colored measurement noise. In the meantime, it alleviates the dynamic model and observation errors that occur during the state parameters estimation. To implement effective vehicle positioning in GPS signal blocked environments, the authors in [30] propose a data fusion method, which integrates multiple complementary low-cost sensors including GPS receiver, MEMS IMU device as well as sliding-mode observer (SMO) sensor. Furthermore, the authors design a federated Kalman filter (FKF) and a hybrid global estimator (HGE) to obtain accurate and reliable vehicle position. However, KF based method is regarded as an optimal filter for linear system with Gaussian noise, while it is not suitable for non-linear systems with non-Gaussian errors.

Alternatively, regression methods and artificial intelligence (AI) methods such as support vector machine for regression (SVR) [31], partial least squares regression (PLSR) [32], input-delayed neural networks (IDNN) [33] as well as back propagation neural networks (BPNN) [34], have been developed for dealing with the positioning issues thanks to their ability to handle the problem of non-linearity and uncertainty [10]. Nevertheless, the above-mentioned approaches suffer from data fluctuation and concept drift during long-term motion dynamics in the city driving especially under GPS outage environments. Specifically, trajectory data collection in urban environments would increase the difficulty of modeling trajectory data and thus degrade the performance of vehicular position prediction in trajectory collection.

In summary, it is essential to the industry seeking an operable method that can realize trajectory data acquisition especially for private cars in large-scale scenarios. Besides, how to precisely obtain vehicle position has become one of the most fundamental challenges for achieving reliable trajectory data acquisition.

B. Contributions

In this study, we first propose a lightweight and user-friendly implementation method for achieving trajectory data acquisition for large-scale private cars. Within the proposed implementation method, we design a tightly-coupled integration of low-cost GPS module and On Board Diagnostics (OBD) reader (see details in Section II). On one side, the vehicle position can be presented by longitude and latitude coordinates obtained from the external GPS receiver; on the other side, the OBD reader is plugged into the vehicle OBD interface to read the information from the in-vehicle motion sensors including velocity and steering direction. Hence no external IMU device is needed. It is noteworthy that the existing works on collection of trajectory data are main for floating cars such taxi and bus, our work is the first effort to provide a feasible and low-cost solution which can obtain large-scale trajectory data and particularly suitable for private cars.

In the urban environments, GPS may fail due to the blockage of GPS signal reception and multipath effect [27], [30]. The motion sensing information retrieved from OBD reader is used to bridge GPS errors. Nevertheless, due to the presence of inherent noises in the GPS and OBD device, trajectory data suffers from data deficiency which leads to trajectory tracking and vehicle position error accumulated over time [35]. These problems not only dramatically degrade the performance of vehicle positioning and trajectory collection but also lower the application value of collected trajectory data especially for fine-grained trajectory big data mining [36].

Therefore, we strive to develop a trajectory prediction approach that addresses the major problems, i.e., non-stationary distribution, concept drift and incremental learning in the private car trajectory data collection under urban environments. To that end, we propose a novel vehicle positioning approach by designing an ensemble learning framework of Gauss process regression (GPR) method to enable reliable trajectory collection and enhance the usability of collected trajectory data. Within the proposed approach, trajectory prediction based on the GPS/OBD integration trajectory data is modeled by GPR.

Besides the advantages of the proposed GPR method that handles noisy and small sample size training, we propose to enhance the generalization ability of GPR model with the

TABLE I
NOTATIONS OF THE MAIN VARIABLES

Parameter	Definition
u	The mean function of a Gaussian process
f(x)	The Gaussian process
N	The size of training samples
y_i	The output, i.e., the predicted position
y_*	The joint posterior distribution of the predicted
$\mu(y_*)$	The joint posterior distribution of the predicted point
$cov(y_*)$	The variance of the prediction in line with the test point
θ	The hyperparameter
P(y|x)	The marginal likelihood function
K	The covariance matrix
X	The explanatory variables
y	The corresponding response variable
S_R	The original regression samples
S_C	The generated classification samples by R2C conversion
ε	The calculation error of a hypothesis
$D_i(X, y)$	A distribution based on the collected trajectory data block
S^t	The series of training datasets
ε_k^t	The calculation error of an individual
A(x)	The acquisition function

continuously-collected trajectory data by realizing incremental learning. To achieve this, we first design a regression to classification (R2C) method, so as to train any ensemble algorithms on nonlinear GPRs. After that, the proposed R2C is integrated with Learn++ to realize the incremental learning addressing the non-stationarity problems during trajectory prediction, wherein a classification-type loss (CTL) function is derived for weights construction in the process of ensemble learning. Furthermore, a Bayes optimization algorithm is devised to realize adaptive parameter tuning in the ensemble-learning based GPR method. The proposed approach implements incremental learning when new trajectory data arrives and hence is able to acquire knowledge in the concept drifting scenarios, in other words, the data fluctuation and error accumulation, which are inevitable because the stability of GPS receiver and motion sensors are limited, can be corrected in a timely manner, thus the OBD reader offers significant supplementary for vehicle positioning when GPS data are of large error in the dense multipath and GPS outage situations.

The main contributions of this paper are outlined as follows.

1) A lightweight and user-friendly implementation method based on GPS/OBD integration, which is able to realize large-scale trajectory data collection for private cars.

2) A trajectory prediction approach via designing an ensemble learning framework of GPR for achieving accurate and reliable trajectory collection.

3) An evaluation of proposed method under various road conditions and GPS outages in real road experiments.

The remainder of this paper is organized as follows. Section II describes the design of GPS/OBD integration device for large-scale private cars trajectory data acquisition. Following that, Section III gives an overview of the proposed vehicle positioning and trajectory prediction approach for trajectory collection. Section IV illustrates the ensemble learning framework with GPR. Section V presents our evaluation of the proposed GPS/OBD integration based trajectory collection method. Section VI concludes this paper. Table I presents the notations of the main variables.

Fig. 1. The GPS/OBD device installed on vehicle.

II. TRAJECTORY COLLECTION VIA GPS/OBD INTEGRATION

A. GPS/OBD Device

Differently from using external IMU device combined with GPS for implementing trajectory collection, our proposed tightly-coupled GPS/OBD integration method utilizes the motion sensor information for the individual vehicle, which is obtained by the OBD reader, to correct errors when GPS signals are outage. Specifically, as shown in Fig. 1, the GPS/OBD integration consists of three components, i.e., GPS module, OBD reader plugging through the OBD interface to read driving information from vehicle motion sensors, the communication unit [37] with a SIM card that is embedded into the GPS module. Note that for the increasing demand of driving safety, anti-lock braking system (ABS) and electronic stability program (ESP) are becoming standard equipment for vehicles manufacturers. Hence, this enables the OBD reader to read driving status including velocity, acceleration as well as driving direction from the in-car motion sensors. We obtain two types of trajectory information, namely the position of vehicle (latitude or longitude), and the driving status, which can be retrieved via the GPS module and OBD reader, respectively. The GPS module can indicate the reception of GPS measurements based on the number of GPS satellites and the strength of received signals. The information on driving status collected from OBD reader contains the trip info, speed ('vel'), the steering direction ('dir'), the revolutions per minute (RPM), and so on. As seen the SIM card in Fig. 1, the collected trajectory data will be modulated and coded in the communication unit by using LTE or 5G technology [38], [39] and then transmitted to the server (see Fig. 2).

B. Discussion on Privacy Concern During Trajectory Collection

The proposed GPS/OBD integration based trajectory collection method poses almost no influence on the road driving and involves negligible privacy issues. As shown in Fig. 2, the GPS/OBD device collects trajectory data and transmits the data back to the trajectory data management (TDM) server through the LTE network.

Specifically, we have taken the following active steps in order to avoid privacy invasion for the private car owners.

i) The International Mobile Equipment Identity (IMEI) number is assigned to the GPS/OBD device as the unique ID for each vehicle and is one-to-one mapped to a bit string as an

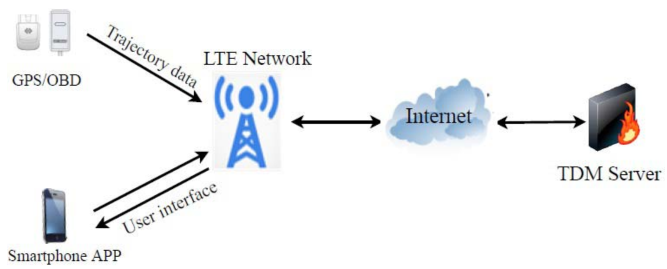

Fig. 2. Trajectory data collection.

anonymized ID for privacy protection. Hence, the trajectory collection method based on GPS/OBD is not to collect any personal information regarding the vehicle (e.g., the brand and the model) as well as the driver or the owner.

ii) The trajectory dataset is stored in the TDM server protected by authentication mechanisms and firewalls in our collaborating company's network. Besides, the TDM server have all the IMEI numbers of the GPS/OBD devices so that is able to identify the collected trajectory data of each vehicle.

iii) We have developed a smartphone App, as shown in Fig. 2, which provides a user interface to access the trajectory data. For the first time to use the App, the private car owners or users need to complete the registration, namely, register their GPS/OBD device via the App to the TDM server by inputting the unique IMEI number. After the authentication from the TDM server, the users are authorized to access and retrieve their own trajectory information from the TDM server via the LTE network. By doing so, the users can easily manage their own trips and fuel consumption, understand the conditions of their cars.

III. Overview of the Proposed Trajectory Prediction Approach

As presented in Section II. A, the trajectory data collected via our device contains rich information including vehicle locations, trip and mileage, driving condition records. Among them, the vehicle locations play a crucial role in the knowledge discovery based on the trajectory data, since any deficiency of vehicle locations would lower the application value of the collected trajectory data. However, the vehicle locations suffer from large error when the GPS outage turns out. The motion sensing information retrieved from OBD reader can be used to bridge GPS errors. To achieve reliable trajectory prediction in GPS outage situation, we propose the trajectory prediction approach based on ensemble learning of GPR method. Within the proposed approach, the trajectory data collected via the GPS/OBD device, which includes vehicle position and motion information, will be mapped into the high-dimensional space based on the kernel function. Then it is able to solve the non-linear and incremental learning during trajectory collection. When the device receives GPS signal properly, we use the trajectory data and train the GPR model. Then the proposed approach can acquire the knowledge that GPS calibrates the accumulation error of motion information from the OBD reader. When GPS signals are outage in some complex urban scenarios, the OBD readings is used to complement the position information so that the GPR model is updated to provide reliable predicted vehicle trajectory.

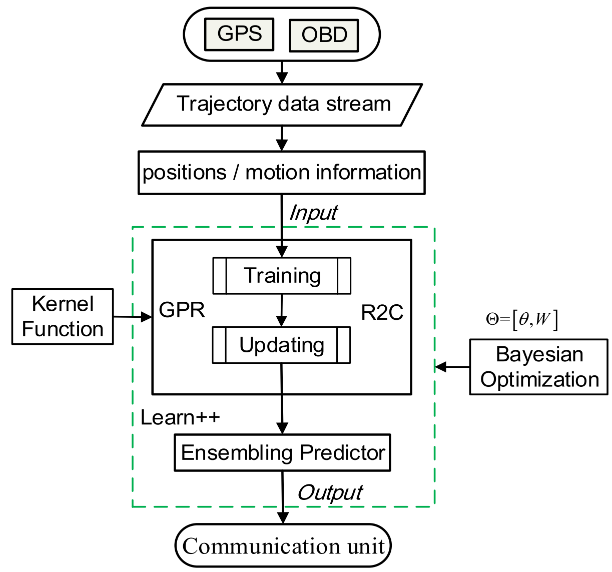

Fig. 3. Proposed trajectory prediction approach.

The structure of proposed approach is illustrated in Fig. 3. We mainly consider the vehicle positions and motion information (such as velocity and steering directions) containing in the trajectory data stream as the input. Note that the trajectory dataset is constantly updated and new trajectory data becomes available within the course of vehicle driving. Due to the inherent instability of the GPS/OBD integration device, such as multipath effect in the GPS signal reception, data fluctuation of on-board motion sensors which partly results from the change of road conditions and driving behavior of drivers in complex urban environments, the underlying distribution of newly-collected trajectory might be different from the previous ones, namely, concept drift problem turns out because of the non-stationary nature [40] in the process of trajectory collection, which in turn affects the training process of GPR method. On the other side, these new trajectory data can be utilized to update the GPR model and incrementally train the model. For this purpose, we design the R2C method to implement ensemble learning with GPR and Learn++, which aims at resolving the non-linearity and non-stationarity in the trajectory data collection. Furthermore, we design a hyperparameter tuning strategy based on Bayesian optimization in order to address the parameter selection for the trajectory prediction approach. Finally, the predicted trajectory data produced by the trajectory prediction approach is output to the communication unit and then transmitted to the TDM server via the LTE network. The details of the proposed approach are presented in the following section.

IV. Ensemble Learning Framework of GPR

A. GPR Model for GPS/OBD Based Trajectory Collection

During the trajectory collection, the heterogeneous information, namely vehicle locations and motion information collected by the external GPS module and on-board sensors, respectively, are put into the GPR model (see *Input* in Fig. 3). GPR is regarded as a flexible and powerful Bayesian non-parametric approach [41], which is to model functions and perform inference on functions. Let u and υ denote the mean function and covariance function of a Gaussian process (GP) $f(x)$, respectively. We have

$$u(x) = E[f(x)], \tag{1}$$
$$\upsilon(x, x') = E[(f(x) - u(x))\ (f(x') - u(x'))], \tag{2}$$

where x and $x' \in R^m$ are random variables. A GP is defined as a distribution over functions,

$$f(x) \sim \mathcal{GP}(u(x), \upsilon(x, x')). \tag{3}$$

The purpose of GPR model is to utilize the sample function, namely training, and characterize the distribution property for the entire value function, in the case of our study, to model, detect and predict GPS/OBD integration based trajectory. Given a training set $\{(x_i, y_i)\}_{i=1}^N = \{X, y\}$, $x_i \in R^m$ denotes the input vector $(id, T_P, lat, lon, vel, dir)$, where id is anonymized bit string as the car identifier, T_P is the time of sampling, (lat, lon) denote the spatial coordinates, vel and dir are the velocity and driving direction. $y_i \in R$ is used to represent the desired output, i.e., the predicted vehicle position. Let N represent the size of training samples.

The process of vehicle positioning and trajectory prediction can be defined as

$$y_i = f(x_i) + n_0, \tag{4}$$

where $\{x_i, y_i, \sigma_i^2 | i = 1, 2, \ldots, N\} = \{\mathbf{x}, \mathbf{y}, n_0\}$ denotes trajectory data. Let $\mathbf{x}, \mathbf{y}, n_0$ denote the vectors of the input, output and output noise, respectively. Based on (4), we wish to estimate the unknown function f from a set of trajectory points $\{\mathbf{x}, \mathbf{y}, n_0\}$. Note that $n_0 \sim \mathcal{N}(0, \sigma_N^2)$, which is uncertain or even zero for the noise-free estimations.

The prior distribution of the output trajectory y can be expressed as follow

$$y \sim \mathcal{N}\left(0, V(X, X) + \sigma_N^2 I_N\right). \tag{5}$$

The joint prior distribution of observed value y and its prediction y_* can be obtained by:

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} V(X, X) + \sigma_N^2 I_N & V(X, x_*) \\ V(x_*, X) & \upsilon(x_*, x_*) \end{bmatrix}\right). \tag{6}$$

Accordingly, the joint posterior distribution of the predicted y_* is given by

$$(y_* | X\ , y, x_*) \sim \mathcal{N}(\mu(y_*), \mathrm{cov}(y_*)), \tag{7}$$

where $\mu(y_*)$ and $\mathrm{cov}(y_*)$ denote the mean and variance of the prediction in line with the test point x_*, respectively. The expressions of $\mu(y_*)$ and $\mathrm{cov}(y_*)$ are given as follows.

$$\mu(y_*) = V(x_*, X)^T \left[V(X, X) + \sigma_N^2 I_N\right]^{-1} y, \tag{8}$$
$$\mathrm{cov}(y_*) = \upsilon(x_*, x_*) - V(x_*, X)^T \left[V\left(X, X + \sigma_N^2 I_N\right)\right]^{-1} V(x_*, X), \tag{9}$$

where $V(X, X) = V_N = (\upsilon_{ij})$ denotes $N \times N$ symmetric positive definite covariance matrix whose element $\upsilon_{ij} = \upsilon(x_i, x_j)$ measures the relativity between x_i and x_j. Note $V(x_*, X) = V(X, x_*)^T$, which represents the $N \times 1$ covariance matrix between x_* and the training inputs X. I_n is an $N \times N$ identity matrix and $\upsilon(x_*, x_*)$ is used to represent the auto-covariance matrix of the test input.

To apply the GPR model into general non-linear regression case, we leverage the kernel smoothing technology to project the input of trajectory collection into high dimensional feature space. The kernel function can enable the original space data to be mapped into high-dimensional space such as to solve the linear inseparable problem. Through the validated experiments, we apply the widely-used kernel function, namely the squared exponential kernel function, which is expressed as follows

$$k(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||_2^2}{2\theta^2}\right). \tag{10}$$

The parameter learned in GPR is hyperparameter, namely θ in (10), which can be obtained by maximizing over the marginal likelihood function $P(\mathbf{y}|\mathbf{x})$ of the GP,

$$\textit{max:} \log P(\mathbf{y}|\mathbf{x}) = -\frac{1}{2}\mathbf{y}^T \Lambda^{-1} \mathbf{y} - \frac{1}{2}\log|\Lambda| - \frac{N}{2}\log(2\pi). \tag{11}$$

In (11), $\Lambda = K(\mathbf{x}, \mathbf{x}) + \sigma_N^2 I_N$. K denotes the covariance matrix, and its entries $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ can be obtained from (10).

B. R2C Ensemble Learning

Despite the advantages of the proposed GPR method that handles noisy and small sample size training for trajectory data, we can further enhance the generalization ability of GPR by using continuously-collected trajectory data to realize incremental learning. To achieve this, we first propose a regression to classification (R2C) method so as to train ensemble algorithms on non-linear GPRs. Then we integrate the proposed R2C into GPR with Learn++ to realize the incremental learning and thus resolve non-stationary problems of trajectory data collection.

We assume that explanatory variables X consists of m dimensional input features. y denotes the corresponding response variable in the trajectory prediction problem. A sample can be split into two subsamples via shifting the value of response variable up and down with a small distance δ. In the R2C process, a regression dataset is transformed into a classification dataset through assigning two different labels 1 and -1, respectively. According to [42], Learn++ is the ensemble of classifiers based approach and learns in the concept-drifting scenarios. However, it fails to be directly applied to regression model since the extra classification error rate computation is required in that case. To resolve this, we construct a classification-type loss (CTL) for R2C to solve the error rate computation in regression problems. As a result, the conversion from regression problems to classification ones is self-contained in the CTL.

The proposed R2C based ensemble learning framework is composed of three steps.

Step 1. We assume the original regression samples $\mathbf{S_R}$ which is denoted by

$$\mathbf{S_R} = \left\{ (X(i), y(i)) \mid X(i) \in R^m, y(i) \in R, i = 1, 2, \ldots \right\}. \tag{12}$$

We can move the sample along the direction of the response variable's dimension with a small δ so that obtain a dataset for classifier training. Therefore, the generated classification samples by R2C conversion is given by:

$$\mathbf{S_C} = \left\{ \begin{array}{l} (X(i), y(i) \pm \delta), \hat{y}(i) \middle| (X(i), y(i)) \in R^{m+1}, \\ \hat{y}(i) \in \{1, -1\}, i = 1, 2, \ldots \end{array} \right\}, \tag{13}$$

where i denotes the index of samples. The samples with $y_i + \delta$ are assigned with label $+1$. The other classifications generated by the same regression samples with $y_i - \delta$ are assigned -1.

The purpose of that we construct CTL function is to enable GPR to act on a virtually-converted binary classification problem. Apparently, the classification hyperplane generated by dataset $\mathbf{S_C}$ is identical with the regression hyperplane generated by $\mathbf{S_R}$.

Step 2. We train an ensemble learning algorithm with GPR by using a base learner on the generated dataset in (13). Based on the GPR model $y_i = f(x_i) + \varepsilon$, for a regression example (x_i, y_i), the corresponding CTL can be defined as follows:

$$[|y_i - f(x_i)|] = \begin{cases} 1, & if\ |y_i - f_i(x_i)| > \delta \\ 0, & if\ |y_i - f_i(x_i)| \leqslant \delta. \end{cases} \tag{14}$$

By making use of the weighting majority voting, the weight of classifier is calculated based on classification error (CE). Here the evaluations of existing classifiers on current trajectory data block is used to compute the CE. After that, the calculation error of a hypothesis can be expressed by

$$\varepsilon = \sum\nolimits_{i=1}^{n} D_i(X, y) [|f(X_i) - y_i| \leqslant \delta], \tag{15}$$

where $D_i(X, y)$ represent a distribution which is obtained based on the collected trajectory data block.

Step 3. We remove the predictors with high error rate during the R2C operation. The regression plane in the process of ensemble learning, which is identical with a plane in the kernel space, is incrementally constructed along with the remaining predictors as well as their weights $\omega \propto 1/\varepsilon$. In other words, predictors in $\{f\}$ are assigned with small weight value when the error on current trajectory block is large.

Specifically, in order to implement R2C ensemble learning, we divide the trajectory data stream into series of training datasets:

$$S^t = \left\{ \left(X_i^t, y_i^t\right) \middle| X_i^t \in X, Y_i^t \in Y, i = 1, 2, \ldots, m^t \right\}. \tag{16}$$

In (16), t is the number of trainings. X denotes the input data, which contains the original regression data blocks. The last column of X represents the response variable of the original regression problems. Y denotes the label sets {1, -1}. m^t is the size of training set at each t.

Accordingly, the calculation error of an individual f_k can be expressed by:

$$\varepsilon_k^t = \sum\nolimits_{i=1}^{m^t} D_i^t\left(X_i^t, y_i^t\right)\left[\left|f_k\left(X_i^t\right) - y_i^t\right| \leqslant \delta\right], \tag{17}$$

where k is the number of existing classifiers and f_k denotes the existing hypothesis (classifier). In each step t, we obtain the distribution $D_i^t(X, y)$ from the currently available data block. Let W denote the size of trajectory data block. By analyzing the *i-th* dataset S^t, namely $\left(X_i^t, y_i^t\right)$, the distribution of D_i^t can be drawn. At $(t+1) - th$ iteration, a new block of data is generated based on the latest D_i^{t+1}. The advantages for the proposed R2C based ensemble learning approach are twofold. i) It is able to adapt the trained models to the current distribution and hence effective in trajectory collection in the non-stationary and concept drifting scenarios; ii) It is incrementally learned and save storage space required for storing historical trajectory data.

It is noteworthy that the concept drift may arise in any iteration and might be occurring continually, the proposed ensemble learning based GPR can learn from trajectory data whose statistical characteristics change during the course of trajectory collection, and thus track it timely and update the model when new trajectory data arrives. As a result, the proposed R2C method combined with Learn++ can be applied to a regression problem so that the proposed approach can solve the non-linearity and non-stationarity and the incremental training problems.

Fig. 4 presents the workflow of the R2C ensemble learning framework of GPR approach. When GPS is available and in a good status, the trajectory data will be directly put into $CACHE_1$ and then the communication unit, in which the data will be stored in $CACHE_2$ and modulated and coded (baseband signal process), and then transmitted through the radio frequency (RF) terminal. When the GPS outage turns out in the current trajectory data, its historical data will be used for training and generating the predicted data. Note that the data collected via our device contains the GPS status, which is determined jointly by the number of GPS satellites and the strength of received GPS signals. In the practical application, we consider two types of GPS outage, namely partial outage and full outage [43]. Partial GPS outage means that in-view satellites are larger than one and smaller than four, full GPS outage indicates that the device is unable to receive any GPS signals, for instance, when the vehicle is driving through a tunnel. To cope with the GPS outage, we use the historical trajectory data to implement GPR training and ensemble learning. As shown in Fig. 4, let [..., x_i, x_{i-1},...,x_{i-N},..., x_2, x_1] denote the continuous trajectory data stream. We assume that there are no GPS outages taking place when we collect data [x_{i-1},..., x_2, x_1]. Hence these data are put into the $CACHE_1$ and then to the communication unit. We detect GPS outage for trajectory data x_i, x_i is put into the ensemble learning module. In the meantime, the training process is initialized. The historical trajectory data of x_i, i.e., [x_{i-1},...,x_{i-N}] is taken as the training dataset and the size

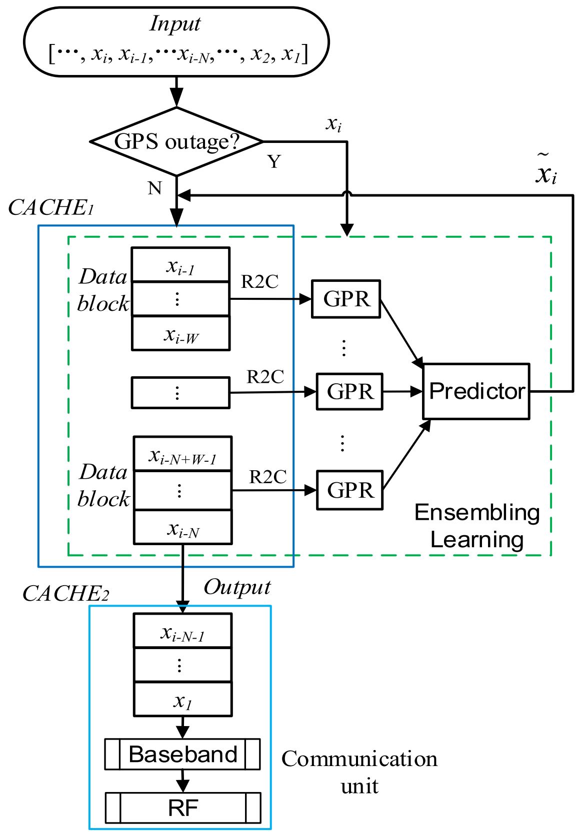

Fig. 4. R2C ensemble learning framework of GPR.

is N. The dataset $[x_{i-1},\ldots,x_{i-N}]$ has been stored in $CACHE_1$ when we collect x_i. Based on (16), the training dataset is divided into several data blocks. In each data block, the R2C based GPR is trained and updated.

Note that the OBD readings included in x_i, particularly the driving information, i.e., the velocity and driving direction, will be used as the input of the GPR for obtaining the prediction of x_i, namely $\tilde{x}_i$. Then $\tilde{x}_i$ will be put back to the sequence in $CACHE_1$. As a result, for the trajectory data stream $[x_i,\ x_{i-1},\ldots,\ x_2,\ x_1]$ and GPS outage takes place in x_i, the sequence $[\tilde{x}_i, x_{i-1},\ldots,\ x_2,\ x_1]$ will be transmitted to the data server via LTE networks. If there is GPS outage in the trajectory data collection after x_i, for example, GPS outage happens in x_{i+1}, the training data for x_{i+1} is $[\tilde{x}_i, x_{i-1},\ldots, x_{i-N+1}]$. In other words, for the trajectory data with GPS outage, on the one hand, we obtain its prediction based our proposed method and send it to the data server; on the other hand, such prediction will be added to the training dataset when continuous GPS outages occur. By doing so, in spite of introducing accumulative errors, the trajectory prediction performance can be guaranteed when the duration of GPS outage is not long.

C. Hyperparameter Selection via Bayesian Optimization

In order to resolve the hyperparameter selection for the proposed trajectory prediction approach, we employ the concept of function optimization method and then design a Bayesian optimization based hyperparameter selection strategy.

We define an unknown objective function $g(\cdot)$ and construct a probabilistic model for $g(\cdot)$. By sampling the objective function the parameters $\Theta = [\theta, W]$ are regarded as independent variables of the function $g(\cdot)$, a prior distribution can be obtained and represents belief on the objective function. Bayesian optimization [44] is an effective method for finding the extrema of the objective functions if they do not have a closed-form expression while can obtain observations by sampling. The posterior distribution of the objective function can be calculated, which is able to capture the updated beliefs about the unknown objective function.

For the sake of flexibility and tractability, we choose the Gaussian process (GP) as prior over objective function $g(\Theta)$. Therefore, the sampling points $\Theta = [\theta, W]$ on unknown function $g(\cdot)$ can be treated as random variables of GP. By selecting parameters Θ, we can assume that these points are part of GP and are subject to multivariate Gaussian distributions [27]. The mean and variance of GP can be calculated based on the properties of the multivariate Gaussian distribution. Without loss of generality, the *priori* mean function is assumed to be zero in this study. For this reason, the GP is completely determined by the covariance functions.

In the Bayesian optimization based hyperparameter selection, the acquisition function is used to construct a utility function and guide the search towards the optimal model parameters $\Theta^* = [\theta, W]^*$. Then the acquisition function is required to maximize so that the next point for evaluating the objective function can be determined. To this end, we wish to sample $g(x)$ at $\mathrm{argmax}_x A(x|S)$, where $A(\cdot)$ is used to denote the acquisition function. We choose Expected Improvement (EI) criteria based acquisition function [45] which can be expressed as

$$A(x) = \begin{cases} (\mu(x) - g(x_{best}))\,\Phi(\gamma(x)) \\ \quad +\sigma(x)\,\varphi(x), & \sigma(x) > 0; \\ 0, & \sigma(x) = 0. \end{cases} \tag{18}$$

In (18), $\mu(x)$ and $\sigma(x)$ denote the predictive mean function and predictive variance function of the objective function $g(x)$, respectively. $x_{best} = \arg\max_{x_i \in x_{1:t}} g(x_i)$ denotes the current best observation. $\varphi(\cdot)$ and $\Phi(\cdot)$ represent the probability density function (PDF) and cumulative distribution function (CDF) of the standard normal distribution, respectively. Here, the root mean square error (RMSE) in trajectory prediction is regarded as a representation of generalization. Combined with Gaussian process and $A(x)$, the optimal combination of $\Theta^* = [\theta, W]^*$ can be found after several iterations and hence the exact and stable ensemble learning based GPR model can be obtained.

Algorithm 1 presents the procedure of the proposed Bayesian optimization based hyperparameter selection. The initial Gaussian process is first obtained by initializing P initial values, namely $I = \{\Theta_{1:P}, r_{1:P}\}$. Then the next set of optimization parameters is obtained and the GPR model is retrained. The posterior Gaussian process is updated by the new set of parameters and the corresponding RMSE r. The optimal combination of parameters will be achieved after several iterations.

The time complexity of Algorithm 1 is $O(TmN^3)$, where T is the number of iterations, m denotes the number of features, N represents the number of training samples. During the practical trajectory data collection, it is not required large N. The proposed trajectory prediction approach can handle the large number of continuous trajectory data stream, since it only needs to store and process a small number of trajectory data (with the order of N). Normally, N is less than 50, i.e., N=40 in the practical application of trajectory collection. Hence, the proposed approach can incrementally learn the newly-collected data in an online manner thereby updating the ensemble learning based GPR method. For these reasons, the computation complexity of trajectory data collection for each individual vehicle is not high and acceptable when applying our proposed approach into the practical large-scale trajectory data collection.

Algorithm 1 Bayesian Optimization Based Parameter Selection

Input: The initial observation $I = \{\Theta_{1:P}, r_{1:P}\}$,where $\Theta = [\theta, W]$.
Initial Settings: Establish training set T_r, testing set T_e.
Output: $\left\{\Theta_t^*, r_t^*\right\}_{t=1}^{T}$.
for t = 1, 2,..., T **do**
  Find x_t by optimizing the acquisition function over the GP: $x_t^* = \arg\max_x A\,(x|I)$.
  Re-train GPR model with parameter $\Theta^* = [\theta, W]^*$ on training set T_r.
  Test GPR model on T_e, obtain the updated RMSE r_t^*.
  Augment the observation set $I = I \cup \left(\Theta_t^*, r_t^*\right)$ and update the GP.
end for

V. Experimental Results

A. Road Experiment for Trajectory Collection

To validate the performance of the proposed trajectory prediction approach, we conduct the road experiments for trajectory data collection on a SUV Ford EDGE 2016 as shown in Fig. 1. The accuracies of motion sensors in the tested vehicle are 0.1 m/s and 10^o for the velocity and steering direction, respectively. The size of training dataset is configured to 40 in the GPS/OBD device and the initial value of block size is 5. Fig. 5 depicts the road test trajectory and GPS outages are considered in various road conditions including straight road, throughway (high speed driving), left-angle turn (low average speed with accelerations/decelerations), and viaduct (continuous turn). Five methods, namely BPNN [34], SVR [31], PLSR [32], kernel ridge regression (KRR) [46] and the FIS-SRG method [27] are employed for comparative study.

B. Analysis of Experimental Results

Fig. 6 illustrates the predicted trajectories in the straight lane driving. The GPS outage in this road section lasts 30s. The reference trajectory (i.e., blue line) records the actual positions when the test vehicle is driving through the straight road, other lines give results which are obtained by applying respectively the proposed approach, FIS-SRG, SVR, KRR, BPNN and PLSR based on the original collected trajectory data. It can be seen that our proposed approach produces more accurate trajectory than those from the comparative methods, since it is much closer to the true trajectory. Fig. 7 presents accumulative errors (AE) of proposed approach and comparative methods during GPS outage. It is shown that AE is growing during GPS outage and the proposed approach is able to keep AE relatively low and thus outperforms the other methods.

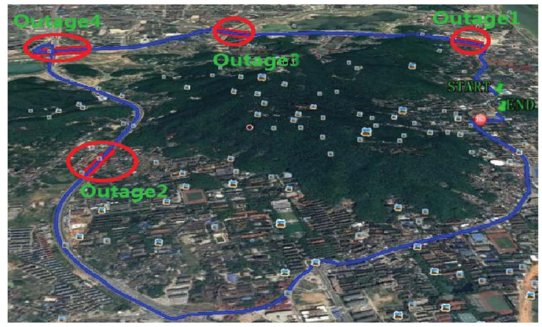

Fig. 5. Experimental trajectory (courtesy of Google Earth).

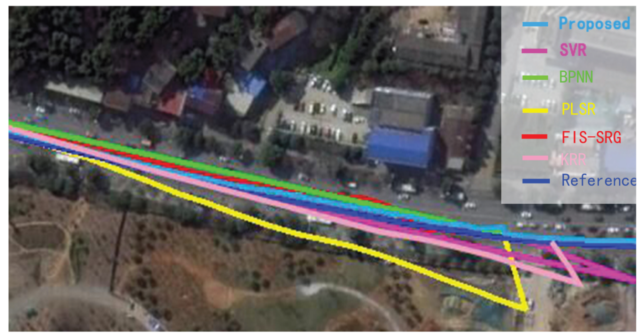

Fig. 6. Predicted trajectories in straight road (GPS outage 1).

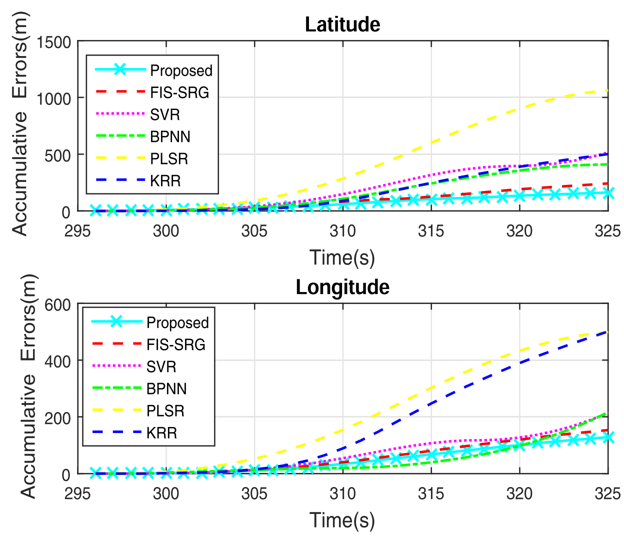

Fig. 7. AE in straight road (GPS outage 1).

Fig. 8. Predicted trajectories in throughway (GPS outage 2).

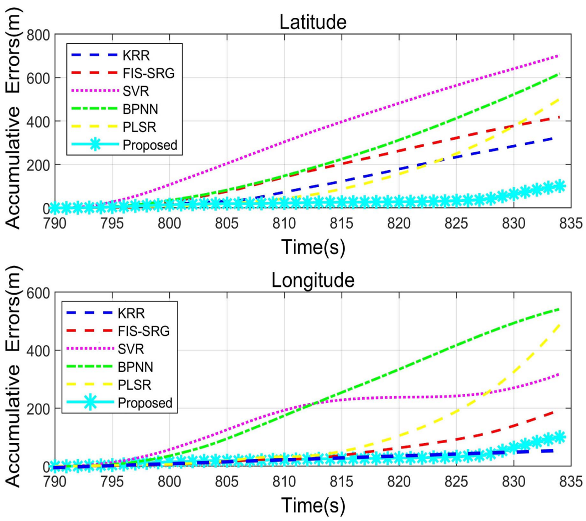

Fig. 9. AE in throughway (GPS outage 2).

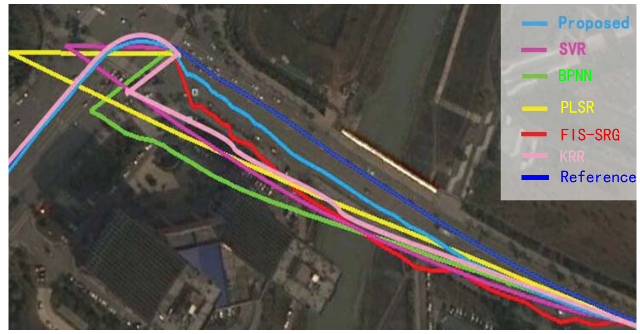

Fig. 10. Predicted trajectories in left-angle turn (GPS outage 3).

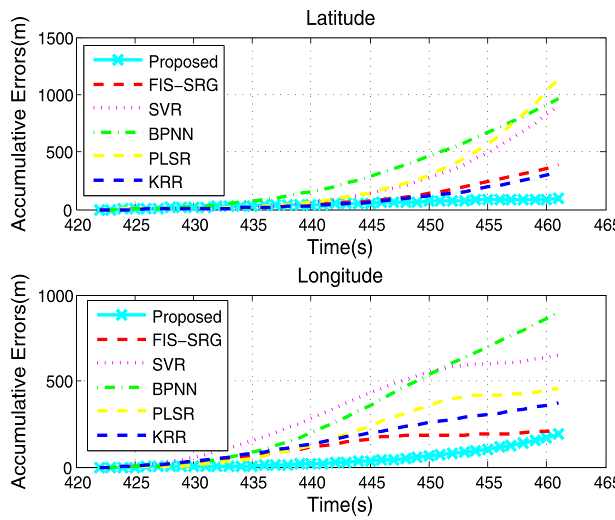

Fig. 11. AE in left-angle turn (GPS outage 3).

The results in Fig. 8 are obtained from the throughway, in which the GPS outage is approximately 45s. The vehicle can drive along the throughway with a relatively high speed, namely 65km/h - 75km/h in our test. In addition, the driving posture of vehicles can be steady when driving through this area. For this reason, all the methods can provide well-matched trajectories. In particular, our proposed method generates near perfect trajectory with the true one because the motion information from OBD reader can solely produce accurate vehicle position under this road condition even GPS outage occurs. Fig. 9 presents the comparison of AE when driving in the throughway. The results validate that the proposed method achieves the lowest AE and thus outperform the comparative methods.

Fig. 10 presents the results obtained from left-angle turn section, in which the GPS outage lasts 40s. The proposed approach generates the best trajectory that nearly fits the true one. However, based on trajectories obtained from FIS-SRG, KRR, SVR, PLSR and BPNN, large trajectory deviations and fluctuations can be seen and the situation becomes worse, namely these trajectory errors are more severe, especially when the duration of GPS outage grows and the drive condition becomes more complex. The reason is that the OBD reader suffers from the concept drift and thus is unable to calibrate the GPS errors when GPS fails during the trajectory collection. It is also supported from Fig. 11, in which the AE during trajectory prediction is illustrated.

When driving to enter the left-angle turn, the vehicle moves with low speed and accelerations/decelerations. Therefore, the information of steering direction from the motion sensors is inaccurate and the output from OBD reader results in azimuth deflection, thus predicted trajectories tend to drift away from the true one, which can be seen in the trajectories from FIS-SRG, SVR, KRR, PLSR and BPNN. For instance, large AE in KRR, PLSR and BPNN results from data fluctuation during trajectory collection in this road condition. Owing to the ensemble and incremental learning ability, the proposed method is capable of mitigating the error accumulation of the OBD data, hence it yields lower position errors both in latitude and longitude than other methods. The above-results are supported from the results in Fig. 11 and Fig. 13. This shows the advantage of the proposed ensemble learning based GPR method dealing with non-stationarity of GPS/OBD data in vehicle positioning and trajectory collection. In a word, the proposed method can obtain more accurate vehicle positions and yield lower AEs both in low-speed and high-speed road conditions when comparing with the baselines.

Fig. 12 presents the results from the viaduct during which the GPS outage is 35s. Under such road condition, the vehicles

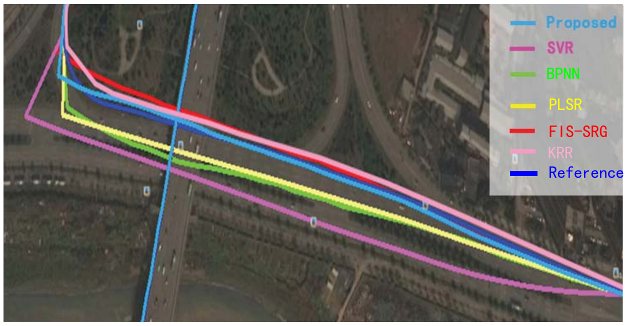

Fig. 12. Predicted trajectories in viaduct (GPS outage 4).

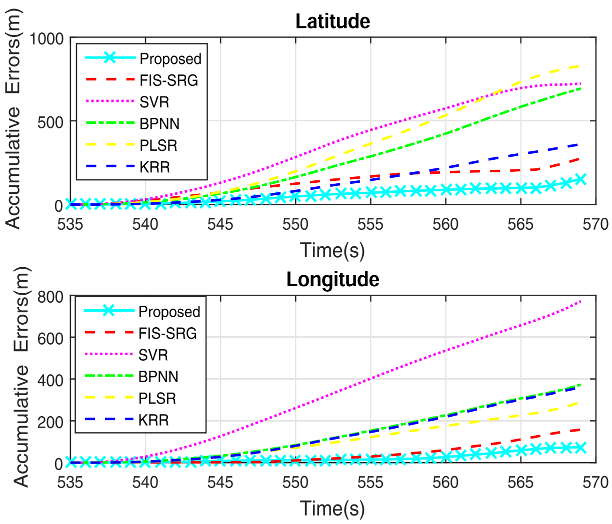

Fig. 13. AE in viaduct (GPS outage 4).

drive with low average speed and keep turning hence the driving maneuvers remain. This significantly affects the performance of vehicle motion sensors. As a result, the motion information obtained from OBD reader is not stable and may not be capable of providing sufficient supplementary for the GPS/OBD integration based vehicle positioning especially when GPS outages turn out. As shown in Fig. 12, there are large trajectory errors when applying SVR, PLSR and BPNN. The proposed approach, KRR and FIS-SRG provide better vehicle positions than other methods, and the proposed approach achieves the best positioning performance and poses less trajectory drift in comparison to all the comparative methods. In addition, the results of AE from Fig. 13 demonstrate that the proposed method only produces small and slow-growth accumulative position errors with the increasing of GPS outage time. The reason behind this is the proposed approach gives a more accurate type-loss, which trains a new classifier and combines dynamically weighted majority voting for the new block of trajectory data. Hence the error is reduced resulting in a more accurate prediction of the updated model.

Moreover, a quantitative comparison of RMSE under four road conditions is presented in Table II. We observe that the proposed approach demonstrates superior performance than the comparative methods under various GPS outages. Based on

TABLE II
RMSE OF TRAJECTORY COLLECTION (m)

Outage	Outage 1	Outage 2	Outage 3	Outage 4
SVR	21.0695	17.0311	33.8094	25.7374
BPNN	13.7896	16.6747	39.5160	18.6677
PLSR	30.8631	18.9983	36.7355	17.2017
FIS-SRG	10.2853	13.6462	23.2273	9.8371
KRR	12.6344	10.5906	25.7443	13.4321
Proposed	**5.9526**	**4.4652**	**10.2607**	**6.2944**

results from left-angle turn (outage 3) and viaduct (outage 4) in Table II, the proposed approach achieves the lowest position errors, i.e., 10m and 6m, respectively. However, long data fluctuation and large trajectory error turn out from other methods, for instance, the RMSE are more than 33m with SVR, BPNN and PLSR from the left-angle turn driving.

VI. CONCLUSION

In this paper, we have presented a low-cost implementation method via designing GPS/OBD integration, which offers a feasible and practical solution to realize trajectory data acquisition particularly applicable for large-scale private cars. In order to enhance the reliability of trajectory collection and ensure the usability of GPS/OBD based trajectory data in complex urban environment, we propose a novel trajectory prediction approach by designing an ensemble learning based GPR method, which is able to cope with the non-linearity, non-stationarity and incremental training problems in trajectory data acquisition. The real-world road experiments validate the feasibility of our approach when performing trajectory collection in various urban environments.

Currently, our method has been successfully applied to real-world trajectory collection and we have obtained a large database of private car trajectory data. With the privacy concern being properly addressed (see Section II. B), we have provided an open access for our private car trajectory data in https://github.com/HunanUniversityZhuXiao/PrivateCarTrajectoryData. Under the consensus on research use only, our trajectory dataset has received wide interests both from the industry and academic. The dataset provides new data source facilitating the emerging ITS applications of trajectory big data analysis in urban computing and smart cities. Note that large volume of data needs to be transmitted during the process of large-scale trajectory data acquisition, hence bringing heavy overhead for the communication and computing of the emerging ITS. Seeking solutions in future work, we will focus on investigating the trajectory data compression and its implementation on our GPS/OBD integration device.

REFERENCES

[1] N. S. Wigginton, J. Fahrenkamp-Uppenbrink, B. Wible, and D. Malakoff, "Cities are the future," *Science*, vol. 352, no. 6288, pp. 904–905, May 2016.
[2] A. Gharaibeh *et al.*, "Smart cities: A survey on data management, security, and enabling technologies," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2456–2501, 4th Quart., 2017.
[3] H. Song, R. Srinivasan, T. Sookoor, and S. Jeschke, *Smart Cities: Foundations, Principles, and Applications*. Hoboken, NJ, USA: Wiley, 2017.

[4] C. N. S. Bureau, *China Statistical Yearbook*. Beijing, China: China Statistical Publishing House, 2019.
[5] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabási, "Returners and explorers dichotomy in human mobility," *Nature Commun.*, vol. 6, no. 1, p. 8166, Nov. 2015.
[6] J. Zhang, G. Lu, H. Yu, Y. Wang, and C. Yang, "Effect of the uncertainty level of vehicle-position information on the stability and safety of the car-following process," *IEEE Trans. Intell. Transp. Syst.*, early access, Dec. 29, 2021, doi: 10.1109/TITS.2020.3044623.
[7] Z. Xiao *et al.*, "Vehicular task offloading via heat-aware MEC cooperation using game-theoretic method," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 2038–2052, Mar. 2020.
[8] Z. Zhou, C. Gao, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, "Social big-data-based content dissemination in internet of vehicles," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 768–777, Feb. 2018.
[9] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 49–55, May 2017.
[10] M. Z. Khan, S. Harous, S. U. Hassan, M. U. G. Khan, R. Iqbal, and S. Mumtaz, "Deep unified model for face recognition based on convolution neural network and edge computing," *IEEE Access*, vol. 7, pp. 72622–72633, 2019.
[11] G. J. Dimitrakopoulos and I. E. Panagiotopoulos, "In-vehicle infotainment systems: Using Bayesian networks to model cognitive selection of music genres," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 8, 2020, doi: 10.1109/TITS.2020.2997003.
[12] V. Cerqueira, L. Moreira-Matias, J. Khiari, and H. van Lint, "On evaluating floating car data quality for knowledge discovery," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3749–3760, Nov. 2018.
[13] F. Zhang, B. Jin, Z. Wang, H. Liu, J. Hu, and L. Zhang, "On geocasting over urban bus-based networks by mining trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1734–1747, Jun. 2016.
[14] W. Yuan, P. Deng, T. Taleb, J. Wan, and C. Bi, "An unlicensed taxi identification model based on big data analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1703–1713, Jun. 2016.
[15] J. Tang, H. Jiang, Z. Li, M. Li, F. Liu, and Y. Wang, "A two-layer model for taxi customer searching behaviors using GPS trajectory data," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 11, pp. 3318–3324, Nov. 2016.
[16] C. Barrios, Y. Motai, and D. Huston, "Trajectory estimations using smartphones," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7901–7910, Dec. 2015.
[17] K.-W. Chen *et al.*, "Vision-based positioning for internet-of-vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 364–376, Feb. 2017.
[18] H. Zhao, C. Wang, Y. Lin, F. Guillemard, S. Geronimi, and F. Aioun, "On-road vehicle trajectory collection and scene-based lane change analysis: Part I," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 206–220, Jan. 2017.
[19] W. Shieh, C. J. Hsu, and T. Wang, "Vehicle positioning and trajectory tracking by infrared signal-direction discrimination for short-range vehicle-to-infrastructure communication systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 368–379, Feb. 2018.
[20] J. Liu, B.-G. Cai, and J. Wang, "Cooperative localization of connected vehicles: Integrating GNSS with DSRC using a robust cubature Kalman filter," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 2111–2125, Aug. 2017.
[21] S. B. Cruz, T. E. Abrudan, Z. Xiao, N. Trigoni, and J. Barros, "Neighbor-aided localization in vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2693–2702, Oct. 2017.
[22] H. Qin, Y. Peng, and W. Zhang, "Vehicles on RFID: Error-cognitive vehicle localization in GPS-less environments," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9943–9957, Nov. 2017.
[23] G. Xie, H. Gao, L. Qian, B. Huang, K. Li, and J. Wang, "Vehicle trajectory prediction by integrating physics- and maneuver-based approaches using interactive multiple models," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5999–6008, Jul. 2018.
[24] X. Wang, X. Jiang, L. Chen, and Y. Wu, "KVLMM: A trajectory prediction method based on a variable-order Markov model with kernel smoothing," *IEEE Access*, vol. 6, pp. 25200–25208, 2018.
[25] Y. Gu, L.-T. Hsu, and S. Kamijo, "GNSS/onboard inertial sensor integration with the aid of 3-D building map for lane-level vehicle self-localization in urban canyon," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4274–4287, Jun. 2016.
[26] J. K. Suhr, J. Jang, D. Min, and H. G. Jung, "Sensor fusion-based low-cost vehicle localization system for complex urban environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1078–1086, May 2017.
[27] V. Havyarimana, D. Hanyurwimfura, P. Nsengiyumva, and Z. Xiao, "A novel hybrid approach based-SRG model for vehicle position prediction in multi-GPS outage conditions," *Inf. Fusion*, vol. 41, pp. 1–8, May 2018.
[28] K. Ansari, "Cooperative position prediction: Beyond vehicle-to-vehicle relative positioning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1121–1130, Mar. 2020.
[29] Y. Jia, S. Li, Y. Qin, and Z. Hou, "New vehicle positioning method based on RMINS/KC and robust adaptive KF," *IEEE Sensors J.*, vol. 18, no. 15, pp. 6319–6326, Aug. 2018.
[30] X. Li and Q. Xu, "A reliable fusion positioning strategy for land vehicles in GPS-denied environments based on low-cost sensors," *IEEE Trans. Ind. Electron.*, vol. 64, no. 4, pp. 3205–3215, Apr. 2017.
[31] D. Bhatt, P. Aggarwal, V. Devabhaktuni, and P. Bhattacharya, "A novel hybrid fusion algorithm to bridge the period of GPS outages using low-cost INS," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2166–2173, Apr. 2014.
[32] B. Li, Y. He, F. Guo, and L. Zuo, "A novel localization algorithm based on Isomap and partial least squares for wireless sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 2, pp. 304–314, Feb. 2013.
[33] A. Noureldin, A. El-Shafie, and M. Bayoumi, "GPS/INS integration utilizing dynamic neural networks for vehicular navigation," *Inf. Fusion*, vol. 12, no. 1, pp. 48–57, Jan. 2011.
[34] N. El-Sheimy, K.-W. Chiang, and A. Noureldin, "The utilization of artificial neural networks for multisensor system integration in navigation and positioning instruments," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 5, pp. 1606–1615, Oct. 2006.
[35] J. Wang, L. Wang, and Y. Song, "Crowd-machine hybrid urban sensing and computing," *Computer*, vol. 54, no. 4, pp. 26–34, Apr. 2021.
[36] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo, "Next-generation big data analytics: State of the art, challenges, and future research topics," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1891–1899, Aug. 2017.
[37] S. Mumtaz, H. Lundqvist, K. M. S. Huq, J. Rodriguez, and A. Radwan, "Smart direct-LTE communication: An energy saving perspective," *Ad Hoc Netw.*, vol. 13, pp. 296–311, Feb. 2014.
[38] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall, "LTE: The evolution of mobile broadband," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 44–51, Apr. 2009.
[39] M. S. Omar *et al.*, "Multiobjective optimization in 5G hybrid networks," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1588–1597, Jun. 2018.
[40] G. Ditzler and R. Polikar, "Semi-supervised learning in nonstationary environments," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2011, pp. 2741–2748.
[41] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (Adaptive Computation and Machine Learning). Cambridge, MA, USA: MIT Press, Jan. 2005.
[42] M. Muhlbaier, A. Topalis, and R. Polikar, "Learn++.MT: A new approach to incremental learning," in *Proc. Int. Workshop Multiple Classifier Syst.*, 2004, pp. 52–61.
[43] V. Havyarimana, Z. Xiao, A. Sibomana, D. Wu, and J. Bai, "A fusion framework based on sparse Gaussian–Wigner prediction for vehicle localization using GDOP of GPS satellites," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 680–689, Feb. 2020.
[44] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Aug. 2012, pp. 2951–2959.
[45] Y. W. Teh, M. Seeger, and M. Jordan, "Semiparametric latent factor models," in *Proc. Artif. Intell. Statist.*, Jan. 2005, pp. 565–568.
[46] J.-W. Kim, I.-H. Kim, and S.-W. Lee, "Decision of braking intensity during simulated driving based on analysis of neural correlates," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2014, pp. 5–8.

Zhu Xiao (Senior Member, IEEE) received the M.S. and Ph.D. degrees in communication and information systems from Xidian University, China, in 2007 and 2009, respectively.

From 2010 to 2012, he was a Research Fellow with the Department of Computer Science and Technology, University of Bedfordshire, U.K. He is currently an Associate Professor with the College of Computer Science and Electronic Engineering, Hunan University, China. His research interests include wireless localization, the Internet of Vehicles, and intelligent transportation systems. He is serving as an Associate Editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Yanxun Chen received the B.S. degree in electronic information engineering from Hunan University of Science and Technology, Xiangtan, China, in 2020. He is currently pursuing the M.S. degree in communication engineering with Hunan University, Changsha, China. His research interests include vehicle positioning, mobile edge computing, and data mining.

Mamoun Alazab (Senior Member, IEEE) received the Ph.D. degree in computer science from the School of Science, Information Technology and Engineering, Federation University Australia. He is currently an Associate Professor with the College of Engineering, IT and Environment, Charles Darwin University, Australia. He is a cyber security researcher and a practitioner with industry and academic experience. He has more than 200 research papers in many international journals and conferences. His research is multidisciplinary that focuses on cyber security and digital forensics of computer systems with a focus on cybercrime detection and prevention. He is the Founding Chair of the IEEE Northern Territory (NT) Subsection.

Hongyang Chen (Senior Member, IEEE) received the B.S. degree from Southwest Jiaotong University, Chengdu, China, in 2003, the M.S. degree from the Institute of Mobile Communications, Southwest Jiaotong University, in 2006, and Ph.D. degree from The University of Tokyo in 2011.

From 2004 to 2006, he was a Research Assistant with the Institute of Computing Technology, Chinese Academy of Science. In 2009, he was a Visiting Researcher at the UCLA Adaptive Systems Laboratory, University of California Los Angeles. From April 2011 to June 2020, he worked as a Researcher at Fujitsu Ltd., Japan. He is currently a Senior Research Expert with Zhejiang Lab, China. He is an Adjunct Professor with Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, China. He has authored or coauthored 100 refereed journals and conference papers. He has been granted or filed more than 50 PCT patents. His research interests include the IoT, data-driven intelligent networking and systems, machine learning, localization, location-based big data, B5G, and statistical signal processing. He was the Symposium Chair or a Special Session Organizer for some flagship conferences, including the IEEE PIMRC, IEEE MILCOM, IEEE GLOBECOM, and IEEE ICC. He is an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL. He has been selected as the Distinguished Lecturer of the IEEE Communication Society from 2021 to 2022.