Dynamic Prediction of Drivers' Personal Routes through Machine Learning

Yue Dai[1], Yuan Ma[1], Qianyi Wang[1], Yi Lu Murphey[1]
Intelligent System Laboratory, University of Michigan-Dearborn
Dearborn, USA

Shiqi Qiu[2], Johannes Kristinsson[2], Jason Meyer[2], Finn Tseng[2], Timothy Feldkamp[2]
Ford Motor Company
Dearborn, USA

Abstract—Personal route prediction (PRP) has attracted much research interest recently because of its technical challenges and broad applications in intelligent vehicle and transportation systems. Traditional navigation systems generate a route for a given origin and destination based on either shortest or fastest route schemes. In practice, different people may very likely take different routes from the same origin to the same destination. Personal route prediction attempts to predict a driver's route based on the knowledge of driver's preferences. In this paper we present an intelligent personal route prediction system, I_PRP, which is built based upon a knowledge base of personal route preference learned from driver's historical trips. The I_PRP contains an intelligent route prediction algorithm based on the first order Markov chain model to predict a driver's intended route for a given pair of origin and destination, and a dynamic route prediction algorithm that has the capability of predicting driver's new route after the driver departs from the predicted route.

I. INTRODUCTION

Personal route prediction (PRP) is an important technology that can be used in a broad range of applications in smart navigation systems, and intelligent vehicles and transportation systems. The commonly used navigation techniques are mostly based on either the shortest or fastest route strategies. However, many drivers do not take shortest nor fastest routes, in particular, when they are on their daily commuting trips between home and work places, shopping and trips in familiar environments. Accurate prediction of personal routes is very useful for a number of applications. With the fast growing technologies in Vehicle to Infrastructure (V2I) and Vehicle to Vehicle (V2V) communications, future vehicles will have the capabilities of receiving real-time contextual information, such as upcoming accident reports, traffic flow and bad road conditions along the drivers' routes, if the routes are known. We envision an intelligent navigation system can incorporate real-time contextual information relevant to the drivers' intended routes to provide drivers with traffic alert information, alternative routes that match a driver's preferences in terms of roadway types and safety requirements. Accurate prediction of personal driving route can also be used to optimize vehicle fuel consumptions [1, 2]. Our previous study showed that if a driver's route can be predicted accurately, then an intelligent system can be developed to predict traveling speed along the chosen route by machine learning from historical trips and real-time traffic congestions [3]. With the knowledge about the traveling speed of individual drivers are then used to effectively optimize vehicle energy consumptions [1,2]. Accurate personal route prediction can also be used to provide better location-based services, e.g. route-specific traffic information and traveling advice can be provided to drivers [4, 5, 6].

Most of the published research works on route prediction are conducted directly on GPS data. This has the disadvantage of processing and storing large amounts of data to present different routes due to the uncertainty of GPS locations [7, 8]. Two trips sharing exactly the same route may have totally different sequences of recorded GPS points. Storing and processing these spatial and temporal trip data could exceed the capacity of current in-vehicle computing systems as more and more trips daily coming in. This big data problem has been addressed in our previous research [8]. Another major challenge in personal route prediction is the uncertainty of driver's intended routes. Although drivers' daily routes show a high degree of temporal and spatial regularity, there is also a certain degree of exceptions. For example, a driver may have a set of origins, destination and routes that are regularly traveled. However someday the driver may need to visit a doctor, pick up a friend at an airport, or take a road trip. Even with a given origin and destination, a driver can take different routes based on traffic flow, weather conditions, road construction, or just mood. In this research focus on route prediction where the origin and destination of a route are given, but the route between the two locations is not and must be predicted.

In this paper, we present an intelligent system, I_PRP, that is developed to accurately predict personal driving route from an origin O to a Destination D at the beginning of the trip, and when a driver deviates from the predicted route, the I_PRP system can quickly learn to re-predict the driver's intended route. A machine learning algorithm has been developed to generate a system of knowledge bases that represent driver's historical route selections. I_PRP uses a probabilistic prediction algorithm for predicting driver's intended driving route based on the system of knowledge bases, and dynamic route prediction algorithm makes a re-prediction of driver's intended route based on the knowledge of driver's historical route selection and the driver's current trip state. Information, such as traffic flow, weather condition, road construction and other parameters are not considered in paper. The paper is organized as follows. Section II gives a brief review of research literatures related to driving route prediction. Section III presents I_PRP system and the three major algorithms developed for learning personal

driving route preference and predicting driver's intended routes at the beginning of the trip and at the point when a driver deviates from the predicted route. Section IV presents the experiments, and Section V concludes the paper.

II. Related Works

In this section we give a brief overview on the algorithms developed for predicting driving routes. In general route prediction techniques can be divided into two categories, closest distance based matching algorithms [7, 9], or systems based on various probabilistic modelling techniques [5, 10, 12, 14]

Froehlich and Krumm proposed a method that predicts a driver's route by using a function of the distance already driven [9]. It used a similarity score based on distances between the two trips to generate clusters of similar trips. Each cluster of trips represents a route. The similarity scores between the on-going trip and each of the existing routes are calculated, and the route that has the highest similarity score is the predicted one. Their system accuracy was close to 20% for the end-to-end prediction, and 40% for the prediction made in the halfway of trip [9]. In [7], a distance-matching based system was presented for predicting the end-to-end route of a vehicle based on location traces of past vehicle trips. The system calculates the "similarity score" between the on-going trip and the historical trips, and the historical trip that has the best match to the beginning of the on-going trip is used as the predicted route.

Probabilistic based methods predict driver's intended route based on a probabilistic distribution among all possible routes from a given origin to a destination, which is either given or predicted. In [13], for a given origin-destination pair, the probability of a link between the origin and destination is a generalized cost function of its distance to the shortest path. Accordingly, all links on the actual shortest path have a link probability of one, and other links between zero and one. Starting from the origin, a repeated random walk procedure adds links successively from node to node with the link selection process at each node governed by the probabilities of the associated next links. At the destination, the route probability corresponds to the product of the associated link probabilities and is used to correct the unequal sampling probability when the resulting route choice set is used for model estimation. In [14], a driver's route prediction is presented based on a probabilistic prediction of the driver's destination. For each candidate destination, the route prediction algorithm planed a route to that destination. Roads on these routes accumulate the probabilities of their respective destinations, giving higher probabilities to roads along the way to higher probability destinations. The algorithm is based on a single parameter that characterizes how efficiently a driver drives. Once this parameter is computed, it does not require storing a history of trips, and it works in places a driver has never visited.

In [5], a hierarchical tree data structure was built to represent all the possible routes from an origin to a destination. Each route is associated a probability which is calculated based on the frequency of the route the driver travelled in the past three years. The historical trips are organized in three different travel time periods. The real-time route prediction algorithm maps the current location of the vehicle's location to the proper location node in the tree and search for the route that has the highest probability from the node to the destination station. When the vehicle deviates from the predicted route, a new prediction is made by finding the route with the highest probability. Markov modelling has been used in driving route prediction.

Simmons et al [10] used a Hidden Markov Model (HMM) for predicting to predicts a driver's intended route and destination through on-line observation of their GPS position during the trip. They further extended the HMM model to incorporate context by augmenting the state representation to with additional factors, such as time-of-day and day-of-week. The approach was evaluated using a corpus of almost a month of real, everyday driving trips. Based on the experiments of ten-fold cross validation over 46 trips, the average accuracy is above 98%. One of the reasons for this high prediction accuracy is that the test data and the training data are from the same corpus and the test data set is small, only four test trips used in each 10-fold validation process.

The personal route prediction methodologies introduced in this paper are innovative. Knowledge about driver's preference in route selection is learned by a machine learning algorithm and the knowledge is represented in multiple probabilistic matrices. The route prediction algorithms have the capability of predicting the driving route before the trip based on the driver's preferences, and dynamically re-predicting driver's intended route after driver deviates from the predicted route. The dynamical prediction is made based on the knowledge of driver's historical trips and the segment of the current trip taken by the driver. The data used in our experiments are recorded over a period of four to five months from two different drivers, which are more extensive than most of the published work in personal route prediction.

III. I_PRP: An Intelligent Personal Route Prediction System

Fig. 1 gives an overview of the I_PRP system. The input to the I_PRP are the GPS coordinates of a driver's trip starting location, "*O*" and the trip destination, "*D*", it makes a prediction of the driver's intended route from *O* to *D* at the beginning of the trip. When a driver deviates from the predicted route, the I_PRP system can quickly learn to re-predict the driver's intended route based on the knowledge learned from driver's historical trips that have the same destination as "*D*". The I_PRP system's route prediction is made based on the knowledge it learned from the driver's historic driving trips. A machine learning algorithm has been developed to generate a system of knowledge bases that represent driver's historical route selections, a personal route prediction algorithm based the Markov chain model has developed to predict personal routes, and a dynamic route prediction algorithm has been developed to accurately re-predict driver's intended route when the driver deviates from the predicted route. These algorithms are described in the following subsections.

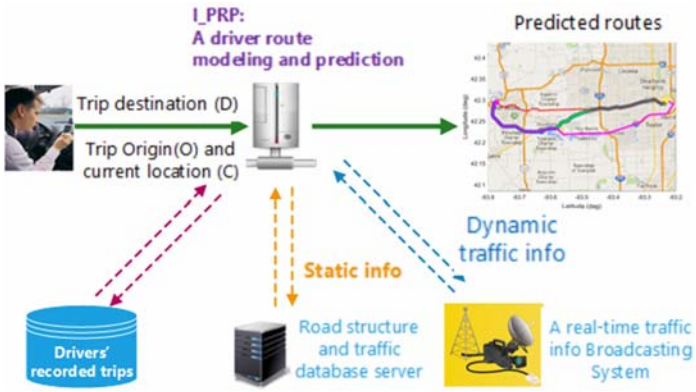


Fig. 1 An intelligent personal route prediction (I_PRP) system

A. A machine learning algorithm for building personal route selecting knowledge bases

The personal driving route prediction problem is formulated as follows. Let Ψ be a set of recorded trips taken by a specific driver. Each trip in Ψ is represented as a sequence of GPS coordinates, Trip = $\{r_1, t_1, r_2, t_2, \ldots, r_i, t_i, \ldots, r_M, t_M \mid r_i = (longitude_i, latitude_i), t_i$ is the time associated with $r_i\}$.

However, trips represented directly in GPS coordinates have the disadvantage of processing and storing large amounts of trip data since trips have identical routes can have different sequences of GPS coordinates. It is more effective in route prediction to use a scheme that maps the trips of the same route to the same representation. In this paper we adopt the canonical route representation proposed in [1]. A trip in the canonical representation is a sequence of link variables, $X = (x_1, x_2, x_3, \ldots, x_N)$, where x_1 is the origin of the trip, i.e. $O = x_1$, x_N is the destination of the trip, i.e. $D = x_N$ and there is only one road that a vehicle can travel from x_i to x_{i+1}, where $i = 1, \ldots, N\text{-}1$. The algorithm that maps a trip of GPS coordinates to the canonical representation can be found in [1]. The Personal Route Knowledge Base (PR_KB) illustrated in Fig. 2 is built from a training data set, Γ, which contains recorded personal trips of the same driver. Two types of knowledge bases have been constructed through machine learning. The first type of knowledge bases is built from trips that have the same origin and same destination (SOSD). The trips in Γ are first partitioned based on their origins and destinations, i.e. $\Gamma = \cup \Gamma_{ij}\ (O_i, D_j)$, $i = 1, \ldots, k_1$, $j = 1, \ldots, k_2$, where Γ_{ij} contains all the trips in Γ that have the same origin (SO) O_i and the same destination (SD) D_j. It is important to point out that the trips in the SOSD partition Γ_{ij} may have many different routes. Fig. 3 illustrates three different routes between a pair of O and D. The trips in Γ_{ij} are used to build a knowledge base, $O_iD_j\_KB$, which consists of a link list Ω_{ij} that contains all the links occurring in the trips in Ω_{ij}, a link transition frequency matrix, LT_F_{ij}, and a link transition probability matrix, LT_P_{ij}, constructed as follows.

Let, $\Omega_{ij} = \{l_1, l_2, l_3, \ldots, l_{Mij}\}$, be a list that contains all possible links occurring in the trips in Γ_{ij}. Both the link transition frequency matrix, LT_F_{ij} and the link transition probability matrix, LT_P_{ij} have the dimension of M_{ij} by M_{ij}, $LT\_F_{ij}\ (l_{h1}, l_{h2})$ is the number of trips in Γ_{ij} in which the driver traveled from link l_{h1} directly to link l_{h2}, where $l_{h1}, l_{h2} \in \Omega_{ij}$, and

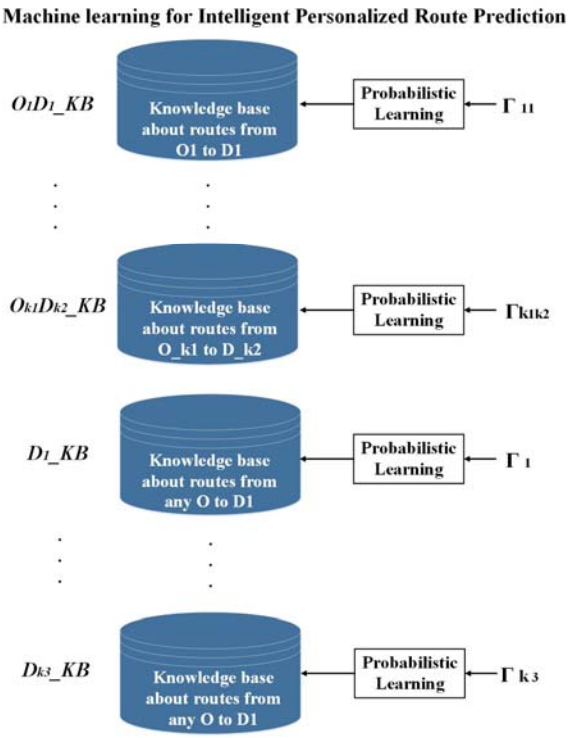


Fig.2 A knowledge base for personal route prediction

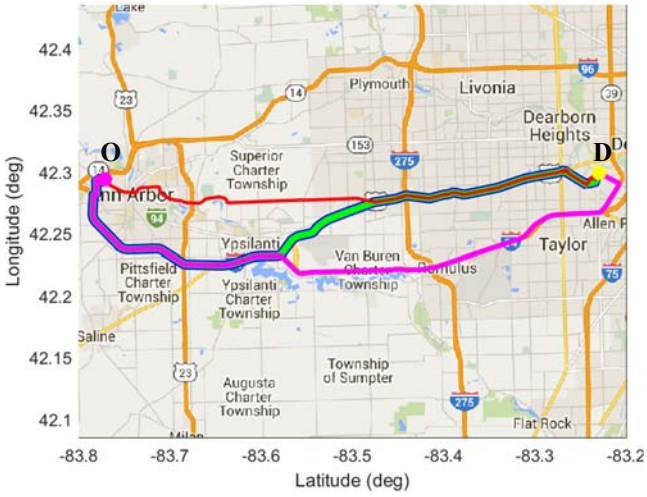


Fig. 3 Illustration of route variations that have the SOSD.

$$LT\_P_{ij}(l_{h1}, l_{h2}) = LT\_F_{ij}(l_{h1}, l_{h2}) / \sum_{k=1}^{M_{ij}} LT\_F_{ij}(l_{h1}, l_k) \quad (1)$$

The second type of knowledge bases are destination knowledge bases, $D_j\_KB = \{D_j, \Omega_j, LT\_F_j$ and $LT\_P_j\}$, for $j = 1, \ldots, k_3$, which are constructed as follows. The training data Γ is partitioned into $\cup \Gamma_j$, $j = 1, ..., k_3$, such that all the trips in Γ_j have the same destination D_j, $j = 1, ..., k_3$, where we assume that there are k_3 different destinations among all the trips in Γ. For each D_j, D_j_KB contains a link list, $\Omega_j = \{l'_1, l'_2, l'_3, \ldots, l'_{Mj}\}$, which includes all possible links occurring in the trips in training data Γ_j. Both the link transition frequency matrix, $LT\_F_j\ (l_{h1}, l_{h2})$, and

the link transition probability matrix LT_P_j (l_{h1}, l_{h2}) have dimensions, M_j by M_j, and LT_F_j (l_{h1}, l_{h2}) is the number of trips in Γ_j in which the driver traveled from link l_{h1} directly to link l_{h2}, where, $l_{h1}, l_{h2} \in \Omega_j$, and

$$LT\_P_j(l_{h1}, l_{h2}) = LT\_F_j(l_{h1}, l_{h2}) / \sum_{k=1}^{M_j} LT\_F_j(l_{h1}, l_k). \quad (2)$$

The machine learning algorithm that builds SOSD and destination knowledge bases is described as follows.

Algorithm 1: Building a knowledge system for intelligent personal route prediction

Input: training data Γ containing recorded personal trips represented in link based canonical form

Output: knowledge system containing knowledge bases $O_iD_j\_KB$, $i = 1, \ldots, k_1, j = 1, \ldots, k_2$, and D_j_KB, for $j = 1, \ldots, k_3$.

Step 1: partition Γ into $\cup \Gamma_{ij}$ (O_i, D_j), such that all the trips in Γ_{ij} have the same origin O_i, and the same destination D_j. Let us assume $i = 1, \ldots, k_1, j = 1, \ldots, k_2$.

Step 2: For each (O_i, D_j) pair, $i = 1, \ldots, k_1, j = 1, \ldots, k_2$,

Step 2.1 extract all the links from the trips in Γ_{ij} and denote them as $\Omega_{ij} = \{l_1, l_2, l_3, \ldots, l_{Mij}\}$

Step 2.2 generate link transition frequency matrix of M_{ij}xM_{ij}, LT_F_{ij} (l_{h1}, l_{h2}) = number of trips in Γ_{ij} that contain two adjacent links, l_{h1} and l_{h2}, such that the driver drove from link l_{h1} directly to l_{h2} in these trips

Step 2.3 use (1) to generate link transition probability matrix of M_{ij}xM_{ij}

Step 3: partition Γ into $\cup \Gamma_j$ (D_j), such that all the trips in Γ_j have the same destination D_j. Let us assume $j = 1, \ldots, k_3$.

Step 4: For each destination D_j, $j = 1, \ldots, k_3$,

Step 4.1 extract all the links from the trips in Γ_j and denote them as $\Omega_j = \{l'_1, l'_2, l'_3, \ldots, l'_{Mj}\}$

Step 4.2 generate link transition frequency matrix of M_{ij}xM_{ij}, LT_F_j (l_{h1}, l_{h2}) = number of trips in Γ_j that contain two adjacent links, l_{h1} and l_{h2}, such that the driver drove from link l_{h1} directly to l_{h2} in these trips, where $l_{h1} \in \Omega_j$ and $l_{h2} \in \Omega_j$.

Step 4.3 use (2) to generate link transition probability matrix of M_jxM_j

Step 5: output $O_iD_j\_KB = \{(O_i, D_j), \Omega_{ij}, LT\_F_{ij}$ and $LT\_P_{ij}\}$, for $i = 1, \ldots, k_1, j = 1, \ldots, k_2$, and $Dj\_KB = \{D_j, \Omega_j, LT\_F_j$ and $LT\_P_j\}$, for $j = 1, \ldots, k_3$.

B. Personal route prediction based on Markov modeling

Let (O, D) be the origin and the destination chosen by a driver at the beginning of the trip. Mathematically, the probability of the driver taking a route can be formulated as a sequence of random link variables, $X = (x_1, x_2, x_3, \ldots, x_N)$, where $O = x_1$, $D = x_N$ and there is a unique road that a vehicle can travel from x_i to x_{i+1}, $i = 1, \ldots, N$-1. A link variable x_i may take one of the M values from the link set $\Omega = \{l_1, l_2, l_3, \ldots, l_M\}$, where Ω contains all possible links from O to D that exist in the training set of driver's recorded historical trips.

The probability of X taking a sequence of links, $L_j = \{c_{j,1}, c_{j,2}, c_{j,3}, c_{j,4}, \ldots, c_{j,N}\}$ is calculated as follows:

$$P(X = L_j | (O, D)) = P(x_1 = c_{j,1}, \ldots, x_N = c_{j,N} | (O, D)) \quad (3)$$

where $c_{j,k} \in \Omega$ ($k = 1, \ldots, N$). Based on the Bayes theorem, we derive the following equation:

$$P(X = L_j | (O, D)) = P(x_N = c_{j,N} | x_{N-1} = c_{j,N-1}, \ldots, x_1 = c_{j,1}) *$$

$$P(x_{N-1} = c_{j,N-1} | x_{N-2} = c_{j,N-2}, \ldots, x_1 = c_{j,1}) * \ldots *$$

$$P(x_2 = c_{j,2} | x_1 = c_{j,1}). \quad (4)$$

For the route prediction problem, we propose to use the following first order Markov Model,

$$P(X = L_j | (O, D)) = P(x_1 = c_{j,1} | (O, D)) *$$

$$\prod_{i=2}^{N} P((x_i = c_{j,i} | x_{i-1} = c_{j,i-1} | (O, D)) \quad (5)$$

Since O and D are given, the values of x_1 and x_N are known, we have $P(x_1 = c_{j,1} | (O, D)) = 1$. Since x_1 takes the value of l_{start} and x_N takes l_{end}, we obtain (6):

$$P(X = L_j | (O, D)) = P(x_N = l_{end} | x_{N-1} = c_{j,N-1}) *$$

$$P(x_2 = c_{j,2} | x_1 = l_{start}) * \prod_{i=3}^{N-1} P(x_i = c_{ji} | x_{i-1} = c_{j,i-1}) \quad (6)$$

The probabilities used in (6) are provided by the link transition probability matrices, LT_P_{ij} (l_{h1}, l_{h2}). Algorithm 2 describes the major computational steps in route prediction.

Algorithm 2 Personal Route Prediction (PRP)

Input: Origin and destination pair (O, D), knowledge bases, O_iD_j, $i = 1, \ldots, k_1, j = 1, \ldots, k_2$.

Output: a most likely route taken by the driver from O to D.

Step 1: finding an OD pair in knowledge base, denote it as O_iD_j for the convenience of description, such that $O_i = O$ and $D_j = D$. Set current link *cnt_link* equal to O.

Step 2: search for a path L from cnt_link to D such that $L = \{l_1, \ldots, l_M\}$, $l_1 = O$, $l_M = D$, and for any other path from O to D, $L' = \{l'_1, \ldots, l'_{M'}\}$, $L \neq L'$, we have

$P(X = L | (O, D)) > P(X = L' | (O, D))$, where

$P(X = L | (O, D)) = \prod_{k=1}^{M-1} LT\_P_{ij}(l_k, l_{k+1})$, and

$P(X = L' | (O, D)) = \prod_{k=1}^{M'-1} LT\_P_{ij}(l'_k, l'_{k+1})$.

Step 3: Output the route L with the highest probability as the predicted route.

C. Dynamic route prediction

Dynamic route prediction is about making re-predictions at the time when a driver took a path that deviates from the current system predicted route. This problem is encountered often by vehicle navigation systems. Many of the systems do not have an effective algorithm to adapt to the drivers' dynamic route changes. We developed a dynamic route prediction algorithm based on the knowledge learned from drivers' route change behaviors to effective re-predict the remaining route in the current trip.

There are many reasons for drivers to take different routes. Based on our study, the most common factors to make drivers to change routes are traffic conditions, traffic light status, road construction and weather conditions. Fig. 4 defines the route change problem within the scope of link-based route representation. The sequence of blue links represents the currently predicted route, the red arrow points to the split point, which is the end of the split link, l_{SP}, and the green link, l_{next}, is the first link that the driver took after the split link, l_{SP}, and the links marked in the brown color are those links the driver took after the l_{next}.

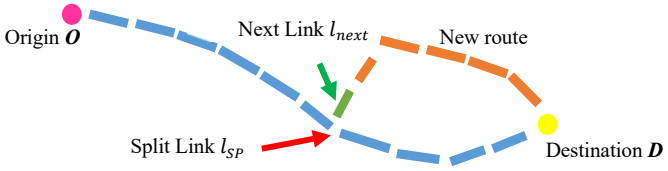
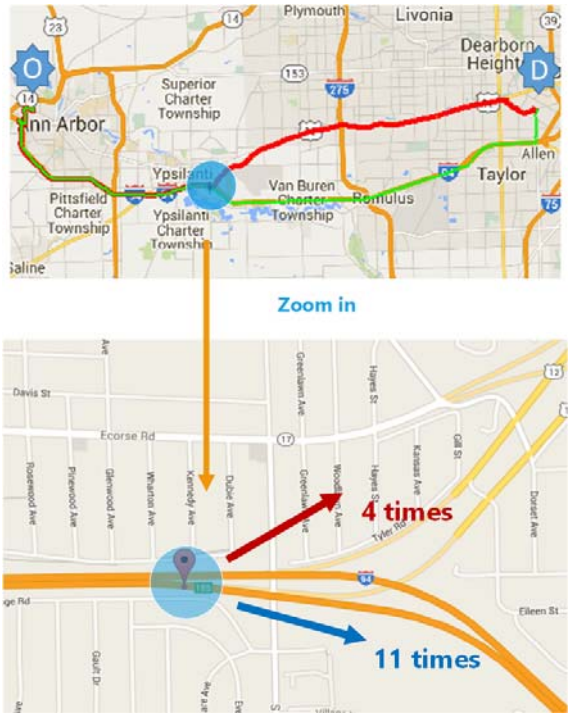


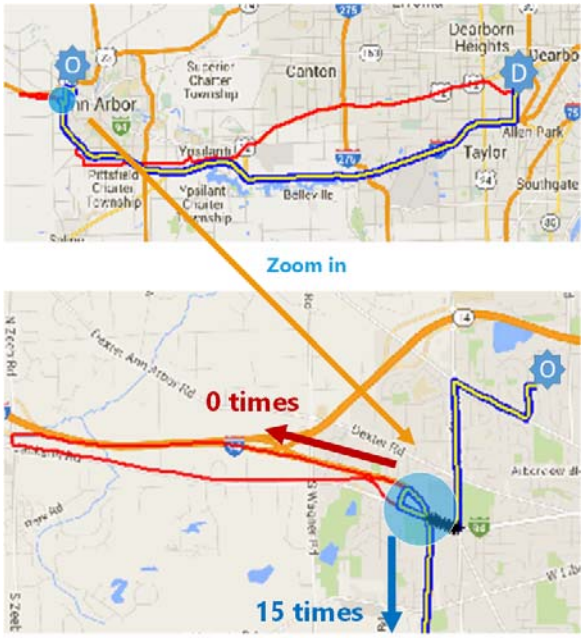
Fig. 4. Illustration of dynamic route change

There are three types of road links at which a driver can change route: highway ramps, highway exit ramps, and intersections in local roads. Fig. 5 (a) and (b) illustrates two of such examples. In these figures, the blue curve is the route predicted by our system at the beginning of the trip, the red curve is the ground truth, i.e. the route the driver took. The sections that show only the curve are where the predicted route and the ground truth are the same. In Fig. 5 (a) shows that the driver split at an intersection, where the predicted route the driver took 11 times and the split route was taken 4 times before by the driver. Fig. 5 (b) the driver split at a local intersection link from a path taken 15 times before. The bottom graph is the zoom-in section circled in red illustrated in the top graph. The zoomed-in graph shows that the driver split from the predicted route (blue curve), which was taken by the same driver 15 times in the past three months, to take a U-turn to avoid a road construction. The route the driver took had not been taken before.

In general, a driver's new route belongs to one of the following four scenarios.



(a)



(b)

Fig. 5. Two examples of driver split routes. In (a) the driver split from a more frequent route to take on a less traveled route. In (b) the driver split from a route taken 15 times before to take a new route.

Scenario 1: the path the driver took at the split point, was taken by the same driver before. One example of this case is shown in Fig. 5 (a), where there was a road construction going on.

Scenario 2: The link after the split point, l_{next} has not been taken by the driver along any route from O to D. However, l_{next} has been taken by the driver on a historical route that has the same destination, D, but a different origin O_i, where $O_i \neq O$.

Scenario 3: The link after the split point, l_{next}, does not belong to any historical route, but it is close to at least one of the historical routes that ends at the same destination as D.

Scenario 4: The link after the split point, l_{next} is far from any of the historical route. In this case, the driver took a new route that has not been taken before.

Based on the analysis of four scenarios, we developed the following dynamic personal route prediction algorithm.

Algorithm 3. Intelligent Dynamic Route Prediction(IDRP)

Input: O and D, the origin and destination of the current trip, currently predicted route, *C_R (t)*, current vehicle location, $l_{current}$, knowledge bases $O_iD_j\_KB$, where $i = 1, \ldots, k_1$, $j = 1, \ldots, k_2$, and a system of destination knowledge bases, D_j_KB, $j = 1, \ldots, k3$.

Output: Predicted route, *C_R(t)*, from the current location to the destination D.

Step 1: If $l_{current} = D$, exit

Step 2: If $l_{current} \neq$ predicted link in *C_R(t)*, then $l_{next} = l_{current}$, otherwise exit

Step 3: If l_{next} is in the knowledge base $O_{h1}D_{h2}\_KB$, where $O_{h1} = O$ and $D_{h2} = D$, $1 \leq h_1 \leq k_1$ and $1 \leq h_2 \leq k_2$.

Step 3.1 set $cnt\_link = l_{next}$ and use step 2 in Algorithm 2 to obtain the route L that has the highest probability that the driver intends to take from l_{next} to D by searching in the link transition probability matrix in $O_{h1}D_{h2}\_KB$

Step 3.2 set *C_R(t)* = L, and exit

Step 4: If l_{next} is in the knowledge base D_{h2}_KB, where $D_{h2} = D$, and $1 \leq h_2 \leq k_3$.

Step 4.1 set $cnt\_link = l_{next}$ and apply step 2 in Algorithm 2 to obtain the route L that has the highest probability that the driver intends to take from l_{next} to D by searching in the link transition probability matrix in D_{h2}_KB

Step 4.2 set *C_R(t)* = L, and exit

Step 5: If l_{next} is close to at least one link within the destination knowledge base D_{h2}_KB, where $D_{h2} = D$, and $1 \leq h_2 \leq k_3$,

Step 5.1: denote the link that is the closest to l_{next} as l_{close}, generate a route that goes from l_{next} to l_{close} using a GIS (Geographic Information System) based on the fastest route. We denote this route as *R_S1*.

Step 5.2: $cnt\_link = l_{close}$ and apply step 2 in Algorithm 2 to obtain the route L that has the highest probability that the driver intends to take from l_{close} to D by searching in the link transition probability matrix in D_{h2}_KB

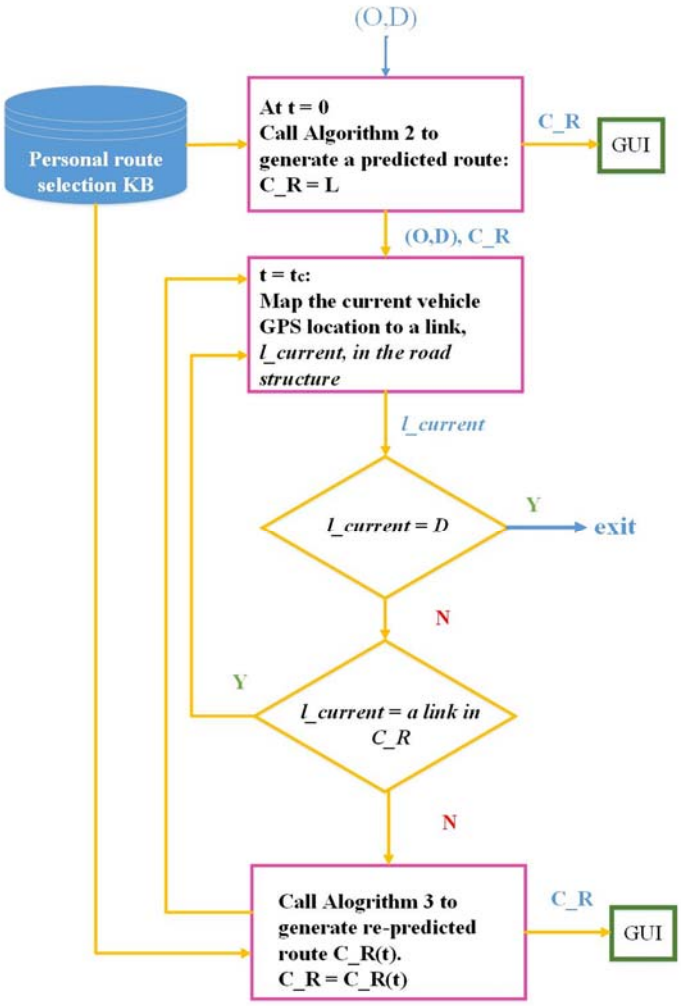


Fig. 6. Computational steps in I_PRP

Step 5.3 set *C_R(t)* = *R_S1* || L, and exit

Step 6: Since there is no knowledge about l_{next}, a route from l_{next} to the destination D can be generated using GIS based on the fastest route, and set this route to *C_R(t)* and exit.

Fig. 6 gives an overview of the computational steps in I_PRP that predicts the driver's intended full route at the beginning of the trip and adjust its prediction when the driver deviates from the current predicted route.

IV. EXPERIMENTS

The Intelligent Personal Route Prediction (I_PRP) system presented in Section III have been evaluated on the real driving data recorded by two drivers, Driver1 and Driver2. For Driver1 we used the 364 trips recorded from January ~ April, 2014 as training data and 22 trips recorded in May, 2014, as test data. In the training data, there are 54 pairs of SOSD and 27 pairs of DOSD, 283 trips belong to the SOSD category and 81 trips belong to DOSD category. In the test data, there are 9 pairs of SOSD, all the trips belong to SOSD category. For all the trips taken by Driver1, in average, each trip contains about 143 links. For Driver2, the 113 trips recorded between March and May

2014 are used as training data and the 19 trips recorded in June 2014 as test data. The training data contain 72 pairs of SOSD and 20 pairs of DOSD, 90 SOSD trips, and 23 DOSD trips. In the test data, there are 10 pairs of SOSD and all the trips belong to the SOSD category. For all the trips taken by Driver2, the average number of links per trip is 223.

For each driver, a system knowledge base presented in Section III is built using the machine learning algorithm, Algorithm 1 from the training data of the driver. Intelligent route prediction algorithms in I_PRP, i.e. Algorithm 2 and 3 are evaluated using the driver's respective test data. The system performance on a test trip is evaluated using the metric defined as follows,

$$Accuracy(T_R) = \sum_{i=1}^{M_2} link\_length(l_i^*) / \sum_{i=1}^{M} link\_length(l_i) \quad (7)$$

where the predicted route generated by I_PRP, $T_R = \{l_1, \ldots, l_M\}$, the true route taken by the driver, $O_R = \{\hat{l_1}, \ldots, \hat{l_{M_1}}\}$, and the link set, $M_L = T_R \cap O_R = \{l_1^*, \ldots, l_{M_2}^*\}$ contains all the links that are predicted correctly by the I_PRP system. The pre-trip prediction performance, i.e. the accuracy of the routes predicted by I_PRP at time t=0 are shown in Table 1. For the purpose of comparison we also applied a shortest route prediction (SRP) program and a fastest route prediction (FRP) program provided by a commercial GIS software to the same test trips and the results are also presented.

Table 1: PREDICTION ACCURACY AT T=0.

	SRP	FRP	I_PRP
Driver 1	21.11%	63.74%	97.38%
Driver 2	57.89%	57.72%	65.76%

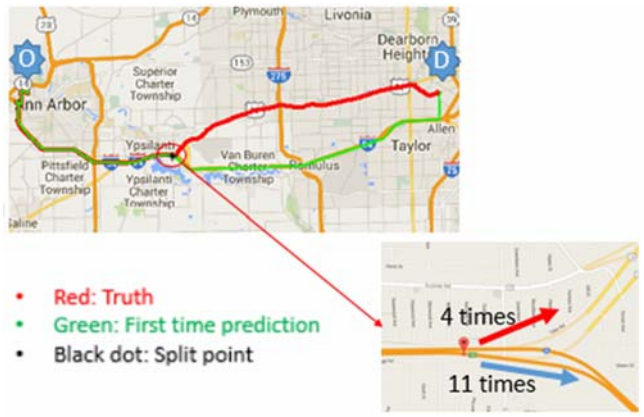
Table 2: Dynamic prediction after splitting.

	A_NRP by SRP system	A_NRP by FRP system	A_NRP by I_PRP
Driver 1	13.76	5.8	1.5
Driver 2	28.64	28.27	3.4

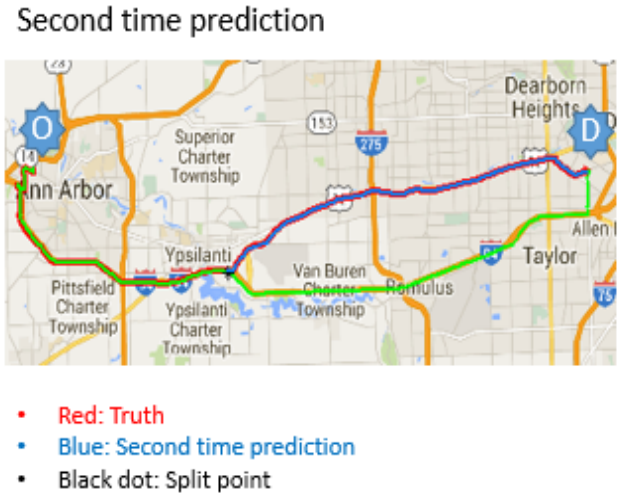
The experiments results generated by the I_PRP is much more accurate than both the SRP and the FRP methods. The results also show that driver 1 did not change route as much as Driver 2. Driver 2 made many dynamic changes during the month of June.

We applied the dynamic prediction algorithm, i.e. Algorithm 3, to the test data of both drivers, and results are summarized in Table 2. The dynamic prediction algorithm is evaluated by using the metric: A_NRP: average number of re-predictions made by the dynamic route prediction algorithm during each test trip, i.e. average number of times that the driver split from the routes predicted by IDRP. If the driver took exactly the same route as predicted by the I_PRP at t=0 on all trips, then A_NRP = 0. The smaller the A_NRP the better the performance of IDRP is.

On the test data of Driver 1, the I_PRP system made 100% accurate prediction over 50% of the test trips at the pre-trip time. The IDRP made one dynamic prediction accurately over 23% of the test trips, and two dynamic predictions over 27% of the test trips. All of the test trips in which the driver split from the



(a)



(b)

Fig. 7 An example of dynamic personal route prediction. (a) shows the route predicted at the beginning, and (b) shows the route predicted at the time after the driver split from the first predicted route.

predicted route were belonging to Scenario type 1 and the average re-prediction time is 1.5 times per trip. When we applied the SRP and the FRP programs to the test data, the average numbers of re-prediction is 13.76 and 5.8 respectively. On the test trips of Driver2, 44% of test trips matched the exact same routes predicted by I_PRP at the pre-trip time, 28% of the test trips belong to the dynamic route selection scenario type 1, and were re-predicted accurately only one time by the IDRP algorithm, 11% were re-predicted twice, and the 17% were predicted more than three times. The average number of re-prediction is 3.4, which is much better than the performances by either SRP or FRP methods, both have the average re-prediction numbers over 28.

Fig. 7 illustrates such an example. In Fig. 7 (a) shows that the driver split from the predicted route (shown in green color) and took a route he took 4 times before, instead of the predicted route, which he took 11 times before. After the split point, the IDRP quickly found the route in the knowledge based and made the correct prediction, as shown in Fig. 7. One of the test trips belong to the dynamic route selection scenario type 2, i.e. the driver took a link that belongs to a historical route he had taken

from a different origin but to the same destination. Based on this knowledge the IDRP algorithm made the correct re-prediction. Within the remaining four test trips, two were re-predicted twice, and the other three trips, the driver took completely different routes than those in the training data. The baseline algorithm has made far more attempts at predicting the personal driving routes than proposed IDRP algorithm.

V. CONCLUSION

We have presented an intelligent system. I_PRP, for personal route prediction, and the three major algorithms employed by the I_PRP, Algorithm 1, the machine learning algorithm for building a personal driving route knowledge system, Algorithm 2, the personal route prediction algorithm based on the first order Markov chain model, and Algorithm 3, the intelligent dynamic route prediction algorithm. We applied the I_PRP system to the trips recorded by two drivers over a period of four months, and our experimental results show that I_PRP has the capability of accurately predict personal driving route from an origin to a destination at the beginning of the trip, and when a driver deviates from the predicted route, the I_PRP system can quickly learn to re-predict the driver's intended route. We also compared the performances of the I_PRP with the performances generated by the SRP and FRP programs provided by a GIS software, the I_PRP has far superior performances. In the future, we will include other conditions such as traffic flow, weather and road constructions in predicting personal driving routes.

REFERENCES

[1] Yoshitaka Deguchi, Kouichi Kuroda, Makoto Shouji, and Taketosh Kawabe, "HEV charge/discharge control system based on navigation information," AE Tech. Pap., vol. No. 2004–2, 2004.

[2] Yi L. Murphey, Jungme Park, Leonidas Kiliaris, Ming Kuang, Abul Masrur, Anthony Phillips, and Qing Wang, "Intelligent Hybrid Vehicle Power Control- Part II: Online Intelligent Energy Management," IEEE Transaction on Vehicular Technology, Vol. 62, No. 1, January, 2013

[3] Jungme Park, Yi L. Murphey, Ryan McGee, Johannes Kristinsson, Ming Kuang, and Anthony Phillips, "Intelligent Trip Modeling for Prediction of Origin-Destination Traveling Speed Profile," IEEE Transaction on Intelligent Transportation systems, Vol. 15, No. 3, June 2014.

[4] Prato, C. G. (2009). Route choice modeling: Past, present and future research directions. Journal of Choice Modelling, 2(1), 65–100.

[5] Karbassi, A., & Barth, M. (2003). Vehicle route prediction and time of arrival estimation techniques for improved transportation system management. In IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No.03TH8683) (pp. 511–516).

[6] Senzaki, D., Chakraborty, G., Mabuchi, H., & Matsuhara, M. (2006). Mobility pattern learning and route prediction based location management in PCS network. In 20th International Conference on Advanced Information Networking and Applications - Volume 1 (AINA'06) (Vol. 2, p. 5 pp.). IEEE.

[7] V. S. Tiwari, S. Chaturvedi, and A. Arya, "Route prediction using trip observations and map matching," Proc. 2013 3rd IEEE Int. Adv. Comput. Conf. IACC 2013, pp. 583–587, 2013

[8] Wang, X., Ma, Y., Di, J., Murphey, Y. L., Qiu, S., Kristinsson, J., Feldkamp, T. (2015). Building efficient probability transition matrix using machine learning from big data for personalized route prediction Procedia Computer Science. Procedia - Procedia Computer Science, 53, 284–291.

[9] J. Froehlich and J. Krumm, "Route Prediction from Trip Observations," Proc. SAE World Congr., vol. 2193, p. 53, 2008.

[10] R. Simmons, B. Browning, Y. Z. Y. Zhang, and V. Sadekar, "Learning to Predict Driver Route and Destination Intent," 2006 IEEE Intell. Transp. Syst. Conf., pp. 127–132, 2006.

[11] Rapoport, A., Gisches, E. J., Daniel, T., & Lindsey, R. (2014). Pre-trip information and route-choice decisions with stochastic travel conditions: Experiment. Transportation Research Part B: Methodological, 68, 154–172, 2014

[12] Frejinger, E., 2007. Random sampling of alternatives in a route choice context. Proceedings of the European Transport Conference, Leeuwenhorst, The Netherlands.

[13] Frejinger, E., Bierlaire, M., Ben-Akiva, M.E., 2009. Sampling of alternatives for route choice modeling. Transportation Research Part B, doi:10.1016/j.trb.2009.03.001 (in press).

[14] Krumm, J., Gruen, R., & Delling, D. (2013). From destination prediction to route prediction. Journal of Location Based Services, 7(2), 98–120.

Acknowledgment

This research is supported by a University Alliance Grant from Ford Motor Company.