Predicting driving behavior using inverse reinforcement learning with multiple reward functions towards environmental diversity

Masamichi Shimosaka[1], Kentaro Nishi[1], Junichi Sato[1], and Hirokatsu Kataoka[1]

Abstract— **Predicting defensive driving is a promising technology for novel advanced driver assistance systems. In recent years, modeling driving behavior in residential roads through inverse reinforcement learning (IRL) has been attracting attention in intelligent vehicle community thanks to the superiority of this approach providing long-term prediction of fine-grained driving behavior. However, it suffers from poor performance in diverse environment due to the fact that the single reward function could not handle all the environment with large diversity. Towards this issue, a novel IRL framework with multiple reward functions to deal with environmental diversity is proposed in the paper. Specifically, the model employs Dirichlet process mixtures as a flexible and powerful Bayesian model to divide the environment into clusters and learns the parameters in each cluster simultaneously. Experimental result with expert driver behavior data shows that our model with multiple reward functions provides superior performance over the IRL model with single reward function. It also suggests that the clustering of environments based on the driving behavior of professional drivers could be useful on evaluating driving environments.**

I. INTRODUCTION

Preventive safety technology has been advancing dramatically in recent years contributing to a reduction in automobile accidents. However, the accident reduction rate is still low on narrow streets in residential areas (residential roads), also known as *Zone 30* [1]. Accidents in residential roads mainly stem from the fact that pedestrians and cyclists are dashing out onto streets while drivers negligent about safety. There has been a lot of researches toward detecting pedestrians [2], [3], however, automatic emergency braking systems based on pedestrian detection are not always reliable on residential roads due to physical limitations such as too short a braking distance. There is therefore a need for preventive safety technologies dealing with such accident factors unique to residential roads.

One form of preventive safety technology that has been proposed for dealing with risk factors on residential roads is pedestrian-to-vehicle (P2V) communication [4]. This type of communication has shown promise as an effective countermeasure to pedestrian accidents on residential roads; however, practical considerations such as the need for constructing an appropriate infrastructure have hindered its implementation.

On residential roads having many uncertainties, a skillful driver would not only observe the legal speed limit but would also practice both *risk anticipation* and *defensive driving*. In other words, a good driver would anticipate potential risks such as pedestrians dashing out onto the street at a blind intersection and would naturally reduce speed beforehand. Consequently, if risk anticipation and defensive driving of a professional driver could be predicted, we could expect it to be applicable to preventive safety technology, that is, to promoting speed reduction by giving warnings to careless drivers and to dealing with risks such as people running out onto the street and inattentiveness to safety. As a pioneering work dealing with this issue, Shimosaka et al. [5], have proposed a technique for modeling the driving behavior of a professional driver on residential roads. They demonstrated the effectiveness of a modeling technique using inverse reinforcement learning (IRL), however, their evaluation where only four courses in the same urban region are used, is severely limited. Since residential roads, the target of prediction, are huge in number, which suggests a highly diverse environment in nature, intensive evaluation with massive driver behavior data and handling the diverse environment are needed. From our observation, the preference on driving through risk anticipation and defensive driving will dramatically change with respect to the conditions such as road width, number of lanes, one-way or two-way street, etc. For this reason, we consider that it would be difficult to predict driving behavior in all environment with the single driver behavior model.

One solution to this problem would be to simply observe the diverse environments and exploit the conditions described above in the model as features. However, it is not completely obvious as to how best to incorporate such environmental characteristics and conditions in the model. In addition, it is sometimes impractical to sense the detailed features of such environmental factors. Even if such massive features could be leveraged, it also raises the concerns about overfitting issues.

Another solution would be to make an individual driver behavior model fitting in each environment. This implies that the model learned individually works successfully even if the used features are limited as in [5]. Nevertheless, there are still issues to aggregate the driver behavior data in each environment, i.e. it is difficult to obtain a sufficient amount of driving data for each course due to the limited human or financial resources. This is because over-training stemming from the poor amount of data obtained in each course could lead the performance degeneration. Moreover, from a practical point of view, there is a need for predicting driving behavior even for unknown courses for which no driving data has been obtained beforehand.

[1]The authors are with the Department of Mechano-Informatics, The University of Tokyo, Tokyo, Japan. e-mail: {simosaka, nishi, jsato, kataoka}@ics.t.u-tokyo.ac.jp

To solve the problem as mentioned above, we provide a flexible way to deal with the environmental diversity and insufficient data by grouping / separating courses having similar / dissimilar driving behaviors and learning the multiple reward functions simultaneously. In this regard, a methodology that improves accuracy when simultaneously learning similar tasks by sharing knowledge among those tasks is generally referred to as multi-task learning, which has been shown to be effective in a variety of applications [6], [7], [8].

Grouping together courses having similar driving behavior and learning the reward function of each group is beneficial not only in improving accuracy but also in classifying courses based on differences in the driving behavior of a professional driver. Taking a similar approach, Straub et al. group roads into clusters based on the driving characteristics obtained for each road by applying the technique of topic modeling to driving behavior [9]. Their technique classifies roads according to the degree of congestion for different time slots based on velocity data obtained through crowd sensing. In our study, we consider that classifying courses according to differences in the way a driver performs risk anticipation and defensive driving could be used in evaluating the potential risks of a road environment.

The rest of this paper is organized as follows: in section II, formalization of driver behavior modeling with a Markov decision process is described. Section III presents proposed framework with multiple reward functions towards environmental diversity. In section IV, we describe the experiments for verifying our model. Finally, section V presents the conclusion of this study.

II. Expressing driving behavior by a Markov decision process

In this section, the formalization of the modeling target and the framework with inverse reinforcement learning is described as a base model of our model.

A. Modeling target

As described in section I, our objective in this study is to predict risk anticipation and defensive driving where a driver anticipates potential risks such as pedestrians dashing out onto residential roads and reduces speed beforehand. To this end, we take the acceleration and deceleration behavior of a driver along a course as the target of modeling. Here, as for what route should be taken to the driver's destination, i.e., the global route is given priori, thus prediction of the destination could be omitted in this paper. This stems from the growing popularity of car navigation technology in recent years. It should be noted that many studies have been provided the way of destination prediction [10], so it is conceivable that we could apply the results of those studies to our research.

On the basis of these assumptions, we treat the portion of linear travel from a left-or-right turn or stop line (start) to the next stop line (goal) as an activity unit and model acceleration/deceleration behavior in that interval (Fig. 1). This can be treated as a route-planning problem from the start to the goal in position-velocity space (Fig. 2).

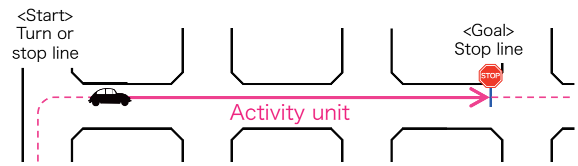

Fig. 1. Assumed situation and activity unit.

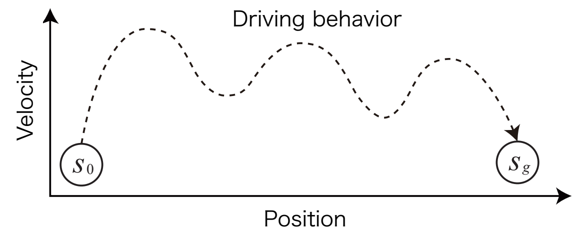

Fig. 2. Target driving behavior.

B. Model formulation

In the base model, route planning in position/velocity space is expressed through an MDP. The underlying graphical model for an MDP is shown in Fig. 3. By using an MDP, driving behavior can be expressed as a state-transition problem where a driver selects certain actions with a goal in mind. Specifically, the dynamics of driving behavior through discrete expressions by denoting state s as a combination of position x and velocity v is expressed, that is, as $s = (x, v)$, discretizing state s together with action a, and defining transition probability $P(s|s', a)$. Moreover, to connect environmental features $\boldsymbol{f}(s) = [f_1(s)...f_D(s)]^\mathsf{T} \leq \mathbf{0}$ that affect driving behavior with reward functions $R(s|\boldsymbol{\theta})$, we make the following assumption.

$$R(s|\boldsymbol{\theta}) = \boldsymbol{\theta}^\mathsf{T} \boldsymbol{f}(s). \qquad (1)$$

The term $f_d(s)$ is described in more detail below. $\boldsymbol{\theta} \geq \mathbf{0}$ denotes the weight parameter.

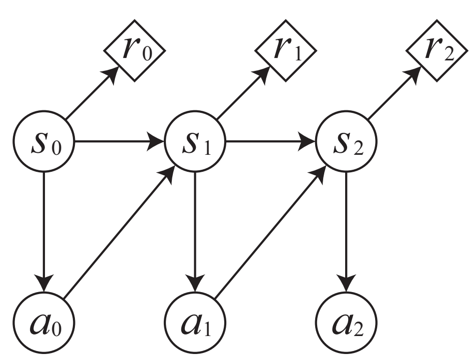

Fig. 3. Underlying graphical model for an MDP.

Likelihood of state series $\zeta = \{(s_0, a_0), (s_1, a_1), ...\}$ and likelihood in data group $\mathcal{D} = \{\zeta_1, \zeta_2, ...\}$ are expressed as follows based on the maximum entropy principle [11].

$$P(\zeta|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(\sum_t (\boldsymbol{\theta}^\mathsf{T} \boldsymbol{f}(s_t) + \log P(s_{t+1}|s_t, a_t))\right)$$

$$P(\mathcal{D}|\boldsymbol{\theta}) = \prod_i P(\zeta_i|\boldsymbol{\theta}), \tag{2}$$

where $Z(\boldsymbol{\theta})$ denotes a normalizing factor.

In the MDP as mentioned above, it is assumed that the driver accelerates and decelerates along the course from the start point to the goal so as to maximize the sum of the reward functions $R(s)$. Here, if reward functions $R(s)$ are given, the state series for maximizing the sum of the rewards obtained from start to goal can be determined by dynamic programming [12].

C. Environmental features

In this paper, we leverage the feature descriptors of state space s proposed by Shimosaka et al. [5] as feature vector $\boldsymbol{f}(s)$ connecting the environmental information onto the reward functions. Specifically, four types of environmental information are used: the start position, the goal position, the position of the near side of an intersection, and the position of the center of the intersection. We decided to use these types of information considering their applicability to the real world and the ease at which they can be obtained (they are included in general maps used for car navigation).

Using the positions of the above environmental factors as reference, we generate potential fields that express five types of behavior in position-velocity space using Gaussian kernels: speed suppression at start/goal points, speed suppression at an unsignalized intersection, suppression of maximum speed, acceleration/deceleration from start to goal, and acceleration/deceleration at an intersection. Here, potential values are taken to be features in each state s. We prepare multiple potential fields that constitute variations in Gaussian-kernel shape (covariance matrix).

III. Inverse Reinforcement Learning with Multiple Reward Functions

In this section, we extend the base model described in the previous section to deal with the environmental diversity. In the base model, the single reward function of IRL [13], [14] can be optimized from the driver behavior data.

In contrast, as described in section I, the technique presented in this paper derives clusters of courses having similar driving behavior and learns a reward function for each cluster with the aim of dealing with environmental diversity and the problem of insufficient data. On the other hand, it is not immediately obvious how to cluster together individual courses and how many clusters should be used, or how that would differ according to environmental diversity and the number of courses for collecting training data. For this reason, our method learns multiple reward functions through IRL based on Dirichlet process mixtures (DPMs) [15]. This approach makes it possible to automatically segment the training data and to estimate the number of the clusters of courses simultaneously. DPMs are known to be sophisticated tools, hence they are often leveraged in the clustering problem from the vectorial data or discrete data. However, the clustering problem on Markov decision processes remains challenge due to the complexity of the models.

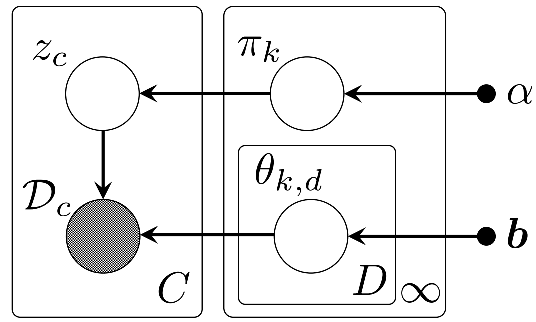

Fig. 4. Graphical model of Dirichlet process mixture inverse reinforcement learning.

The technique most related to our framework is that of Choi et al. [16]. Their technique uses the Dirichlet process prior distribution in IRL and performs efficient optimization by the Metropolis-Hastings (MH) algorithm. When multiple state series are given as training data, the technique handles each separately and learns how to allocate them to clusters. In the framework of our study, however, it is assumed that the data obtained from identical courses have been obtained from the same reward function and that such data is to be optimized by making them belong to the same cluster in the learning process.

A. Optimization of multiple reward functions

We assume that training data consists of C courses. Let M_c $(c = 1, ..., C)$ be the number of trips in each course. Let $\mathcal{D}_c = \{\tilde{\zeta}_{c,1}, ..., \tilde{\zeta}_{c,M_c}\}$ be the state space sequences in course c and $\mathcal{D} = \{\mathcal{D}_1, ..., \mathcal{D}_C\}$ be the overall collection of state space sequences. In addition, driving-behavior data $\tilde{\zeta}_{c,i}$ denotes a series consisting of state s and action a expressed as $\tilde{\zeta}_{c,i} = \{(\tilde{s}_{c,i,0}, \tilde{a}_{c,i,0}), ..., (\tilde{s}_{c,i,T_{c,i}}, \tilde{a}_{c,i,T_{c,i}})\}$. Here, $T_{c,i}$ denotes the number of frames of state series $\tilde{\zeta}_{c,i}$.

Our method divides these data into a number of clusters using the Dirichlet process mixtures and learns multiple reward functions. The weight vector of reward function in c-th cluster is written as $\boldsymbol{\theta}_k$. Additionally, $z_c = k$ expresses the assignment variable of course c to cluster k, which can be written as $\boldsymbol{z} = \{z_c\}$. The graphical model of IRL using the Dirichlet process mixtures is shown in Fig. 4. Here, α denotes a hyperparameter of the Dirichlet process and $\boldsymbol{b}$ is a parameter of the prior distribution $P(\boldsymbol{\theta}_k|\boldsymbol{b})$ of reward-function weight $\boldsymbol{\theta}_k$. The prior distribution is explained in section IV. The joint posterior distribution of the model parameters is defined as follows.

$$P(\boldsymbol{z}, \{\boldsymbol{\theta}_k\}|\mathcal{D}, \alpha) = P(\boldsymbol{z}|\alpha) \prod_k \prod_{c' \in \{c|z_c = k\}} P(\boldsymbol{\theta}_k|\mathcal{D}_{c'}) \tag{3}$$

According to Bayes theorem, the posterior $P(\boldsymbol{\theta}_k|\mathcal{D}_c) \propto P(\mathcal{D}_c|\boldsymbol{\theta}_k)P(\boldsymbol{\theta}_k|\boldsymbol{b})$ could be factorized, where $P(\mathcal{D}_c|\boldsymbol{\theta}_k)$ is given by (2).

Inference is performed using the MH algorithm and by alternately sampling $\boldsymbol{z}$ and $\{\boldsymbol{\theta}_k\}$. First, z_c is updated by sampling from the following.

$$P(z_c|\boldsymbol{z}_{-c}, \{\boldsymbol{\theta}_k\}, \mathcal{D}, \alpha) \propto P(\mathcal{D}_c|\boldsymbol{\theta}_{z_c})P(z_c|\boldsymbol{z}_{-c}, \alpha), \quad (4)$$

where $P(z_c|\boldsymbol{z}_{-c}, \alpha)$ is expressed by the Chinese restaurant process and $\boldsymbol{z}_{-c} = \{z_i|i \neq c \text{ for } i = 1, ..., C\}$ as follows.

$$P(z_c|\boldsymbol{z}_{-c}, \alpha) \propto \begin{cases} n_{-c,z_c}, & \text{if } z_c = z_j \text{ for some } j \\ \alpha, & \text{if } z_c \neq z_j \text{ for all } j \end{cases} \quad (5)$$

Here, $n_{-c,z_j} = |\{c_i = c_j|i \neq c \text{ for } i = 1, ..., C\}|$.

Next, $\boldsymbol{\theta}_k$ is sampled from

$$P(\boldsymbol{\theta}_k|\boldsymbol{z}, \mathcal{D}) \propto P(\boldsymbol{\theta}_k|\boldsymbol{b}) \prod_{c' \in \{c|z_c=k\}} P(\mathcal{D}_{c'}|\boldsymbol{\theta}_k). \quad (6)$$

Due to the fact that it is impossible to draw a sample directly from the distribution $P(\mathcal{D}_c|\boldsymbol{\theta}_k)$, MH algorithm are employed as a sampler of $P(\mathcal{D}_c|\boldsymbol{\theta}_k)$.

B. Prediction in unknown environments

Predicting driving behavior on an unknown course requires that an appropriate decision be made as to which cluster should be used from those already obtained. Here, we assume that data from one or more runs has already been obtained as evaluation data on the course targeted for prediction, and using this data, we select the model with the highest likelihood.

However, taking real-world applications into account, we must also consider cases that are not included in training data and for which evaluation data could not be obtained. In these cases, predicting driving behavior on an unknown course would require the use of environmental characteristics to select which model to use from those previously learned or to determine what weight to apply to a model. In this regard, the field of crowd sensing has been quite active in recent years and methods for obtaining a variety of road characteristics from a crowd have been proposed [17]. Model selection using road characteristics in this way is left as a future topic of study.

IV. Experiment on predicting driving behavior

We performed an experiment to demonstrate the effectiveness of predicting driving behavior by IRL consisting of multiple reward functions taking environmental diversity into account.

A. Experimental data

In the experiment, we recruited one expert driver from a taxi firm and collected data of driving behavior. We recorded the location and speed of the driver behavior from the vehicular network. In addition, we manually added environmental factors as annotations after the experiment based on data obtained from a LIDAR system installed on the vehicle used in the experiment (see Fig. 5).

We selected seven courses in total from two areas in Tokyo, Japan. In each course, it begins from a start line or left / right turn and ends with a stop line. These courses with background maps are shown in Fig. 6. Additionally the course information and number of driving data obtained in the experiment are listed in Table I. It should be noted that each course includes several unsignalized four-way intersections whereas the distance between intersections, width of crossroads, etc. are different with respect to the courses. Furthermore, when traveling along the same road in the opposite direction, that trip is taken to be a separate course since that visibility at intersections and other characteristics will probably differ. As shown in Fig. 6, courses 2 and 3 and courses 5 and 6 are such courses representing travel on the same road but in the opposite direction. Additionally, with the aim of modeling driving that anticipates potential risks on residential roads, we have excluded data on dynamic environmental changes such as the appearance of pedestrians that could affect driving behavior.

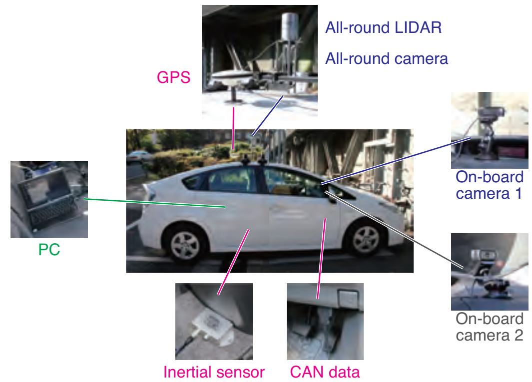

Fig. 5. Vehicle and the installed sensors used in the experiment

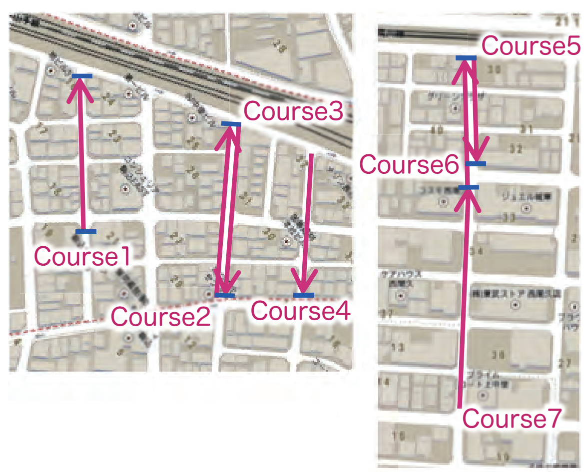

Fig. 6. Experimental courses. The blue lines indicate stop lines. The background maps are cited from Google maps[18].

B. Experimental settings

In this study, as described in section II, we formulated the route-prediction problem in position-velocity space by MDP in discrete state space. Borrowing from the procedure presented in [5], we discretized velocity into 0.5 m/s intervals in 17 steps ranging from 0.5 m/s $= 1.8$ km/h to 8.5 m/s $= 30.6$ km/h. This range of velocities covers speeds slower than 4.0 km/h—the speed of a human being

TABLE I

INFORMATION OF EXPERIMENTAL COURSES AND OBTAINED DRIVING DATA.

Course ID	Course length	# of data
Course 1	90 m	7
Course 2	95 m	6
Course 3	95 m	9
Course 4	79 m	10
Course 5	65 m	9
Course 6	70 m	10
Course 7	135 m	4

taking a leisurely walk—as well as the legal speed limit of $30.0\ \mathrm{km/h}$. In addition, we discretized time into 5 Hz intervals considering that a human being takes about one second to brake after recognizing some sort of danger and that predictions should be performed with even finer granularity. Based on the above discretization, the distance covered by the vehicle in 0.2 s steps range from $0.5\ \mathrm{m/s} \times 0.2\ \mathrm{s} = 0.1\ \mathrm{m}$ to $8.5\ \mathrm{m/s} \times 0.2\ \mathrm{s} = 1.7\ \mathrm{m}$.

We also discretized behavior into accelerate, maintain (speed), and decelerate. Thanks to leveraging the discretization mentioned above, the velocity and the position as integer values in the form of $v_d = 1, ...17$ and $x_d = 1, ...$, the current state can be expressed as $s_t = (x_d, v_d)$, as shown in Fig. 7. The next state can then be expressed as $s_{t+1} = (x_d + v_d + 1, v_d + 1)$ if accelerating, $s_{t+1} = (x_d + v_d, v_d)$ if maintaining speed, and $s_{t+1} = (x_d + v_d - 1, v_d - 1)$ if decelerating. In other words, if we specify the minimum unit of distance to be 1, the distance covered will range from 1-17 per step.

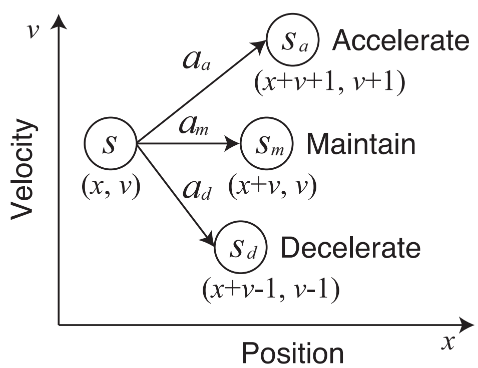

Fig. 7. State and action representation.

Furthermore, for the same environmental factors, we generated multiple features by varying the width of Gaussian kernels, and since we could expect the weights of most potential fields to be 0, we used the Laplace distribution with an average of 0 for the prior distribution of weights. We determined the other parameters through cross-validation.

C. Evaluation metrics

We use the modified Hausdorff distance (MHD) [19] in distance and velocity space to evaluate the degree to which the state series in actual driving matches the state series generated using learned reward functions. The MHD extends the Hausdorff distance to match up data having time-series properties. Given point sets $P = \{p_t\}_{0 \le t < T_p}$ and $Q = \{q_t\}_{0 \le t < T_q}$ having time-series properties, the distance between P and Q can be defined by the following equation.

$$h_\beta(P, Q) = \mathrm{ord}^\beta_{p \in P} \left(\min_{q \in N(C(p))} d(p, q) \right). \tag{7}$$

Here, $N(q)$ denotes the set of points near point q within point sequence Q and $C(p)$ denotes point q within point set Q related to p within point set P. In addition, $\mathrm{ord}^\beta_{p \in P} f(p)$ is the value of $f(p)$ among those calculated for all points within point sequence P for which the percentile of that sequence arranged in ascending order is β. The above measure, however, is a directed one, so in this study, we perform our evaluation by calculating an undirected measure using $H_\beta(P, Q) = \max(h_\beta(P, Q), h_\beta(Q, P))$. Specifically, we first calculate the MHD between state series P in actual driving behavior and 100 state series generated by sampling from the start point using obtained reward functions and then take the average of the MHD values so obtained to perform our evaluation.

Parameter β of MHD is set to $\beta = 0.5, 0.9$. We point out here that $\beta = 0.5$ represents the median of the MHD between the two point sets while $\beta = 0.9$ represents the 90 percentile of the MHD with values arranged in ascending order. In the following, the former is referred to as MHD_{50} and the latter as MHD_{90}.

In the evaluation, we used one of the seven courses as a test course and the remaining six as training courses to learn models. For the test course, we used one run of driving-behavior data as test data and attempted to fit remaining driving-behavior data to obtained models. The model with the maximum likelihood was used in the evaluation. We applied cross-validation to all the combinations, and then obtained the average as an evaluation criterion.

D. Experimental results

Table II shows the performance comparison of our proposed that learns multiple reward functions by grouping all data into clusters based on driving behavior method with the existing technique that learns a single reward function using all data. As shown in the table, the proposed technique exhibits superior performance for both MHD_{50} and MHD_{90}. These experimental results were found to have a significant difference at a significance level of 5% by the t-test. It was also found that courses were assigned to 3.14 clusters on average when performing learning from different combinations of six courses out of a total of seven courses.

To qualitatively interpret the manner in which models are separated, we performed additional experiment where all the seven courses are used as training data. The results of assigning these courses to clusters are given in Table III and are shown in Fig. 8. In general, it is difficult to discuss the validity of clustering results, but it makes sense that course set 2 and 3 and course set 5 and 6 would each be assigned to the same cluster since each represents courses on the same road but running in the opposite direction. Although

TABLE II
EXPERIMENTAL RESULTS WITH MHD_{50} (MEDIAN OF MHD) AND MHD_{90} (90 PERCENTILE OF MHD).

Method	MHD_{50}	MHD_{90}
Single reward function	1.016 ± 0.01510	2.176 ± 0.2183
Proposed	$\mathbf{0.9737 \pm 0.04295}$	$\mathbf{1.943 \pm 0.1940}$

visibility at intersections differs with respect to the direction, factors such as road width, distance between intersections, and number of pedestrians would remain the same, thus the resulting driver behaviors are similar. We consider that clustering courses based on trends in the driving behavior of a professional driver in this way is essentially classification based on course hazards that are potentially indicated by differences in that driving behavior. In short, this clustering approach could also be used in evaluating road environments.

TABLE III
CLUSTER ASSIGNMENTS OF EXPERIMENTAL COURSES.

Course ID	1	2	3	4	5	6	7
Cluster ID	1	1	1	2	3	3	2

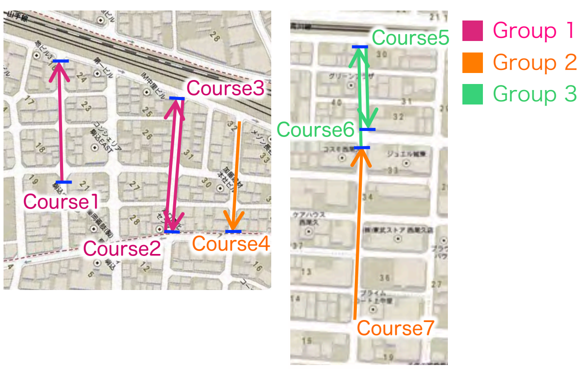

Fig. 8. Clustering results. The background maps are cited from Google maps[18].

V. CONCLUSION

This paper presented a novel driver behavior model on residential roads by using inverse reinforce learning with multiple reward functions in order to avoid insufficient amount of training data on individual courses while taking environmental diversity into account. The technique proposed in the paper extends a maximum-entropy inverse reinforcement learning employing Dirichlet process mixtures, which makes it possible to automatically infer the number of reward functions. This paper also showed that clustering of environments based on the driving behavior of professional drivers could be useful on evaluating driving environments.

As future work, we have to extend our framework that can automatically choose the appropriate cluster where no driving data has been obtained. This type of model selection would require environmental information other than the driving behavior of a professional driver, so the relationship between cluster assignment and environmental information will have to be further studied. There is also the possibility that the driving actions selected by a professional driver will change not only because of environmental differences between courses but also because of weather and time of day. Accordingly, supporting diversity in factors other than environmental ones is also a future topic of the study.

ACKNOWLEDGMENT

We would like to sincerely thank Mr. Keita Sato and Mr. Toshinobu Fukukura of DENSO Corporation for their constructive comments and helpful suggestions for applications of driver behavior prediction in commercial use.

REFERENCES

[1] Institute for Traffic Accident Research and Data Analysis. *ITARDA INFORMATION (in Japanese)*, 2013.
[2] Hyunggi Cho, P.E. Rybski, A. Bar-Hillel, and Wende Zhang. Real-time pedestrian detection with deformable part models. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pp. 1035–1042, June 2012.
[3] S. Silberstein, D. Levi, V. Kogan, and R. Gazit. Vision-based pedestrian detection for rear-view cameras. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pp. 853–860, June 2014.
[4] Martin Liebner, et al. Active safety for vulnerable road users based on smartphone position data. In *Proc. of IV*, pp. 256–261, 2013.
[5] Masamichi Shimosaka, Takuhiro Kaneko, and Kentaro Nishi. Modeling risk anticipation and defensive driving on residential roads with inverse reinforcement learning. In *Proc. of ITSC*, 2014.
[6] Rich Caruana. Multitask learning. *Machine Learning*, Vol. 28, No. 1, pp. 41–75, 1997.
[7] Masamichi Shimosaka, Takahito Ishino, Hiroshi Noguchi, Tomomasa Sato, and Taketoshi Mori. Detecting human activity profiles with Dirichlet enhanced inhomogeneous Poisson processes. In *Proc. of ICPR*, pp. 4384–4387, 2010.
[8] Shoji Tominaga, Masamichi Shimosaka, Rui Fukui, and Tomomasa Sato. A unified framework for modeling and predicting going-out behavior. In *Proc. of Pervasive*, pp. 73–90, 2012.
[9] Julian Straub, Sue Zheng, and John W Fisher Iii. Bayesian nonparametric modeling of driver behavior. In *Proc. of IV*, pp. 932–938, 2014.
[10] Brian D. Ziebart, Andrew L. Maas, Anind K. Dey, and J. Andrew Bagnell. Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In *Proc. of UbiComp*, pp. 322–331, 2008.
[11] Brian D. Ziebart, et al. Maximum entropy inverse reinforcement learning. In *Proc. of AAAI*, pp. 1433–1438, 2008.
[12] Martin L. Puterman. *Markov Decision Process: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.
[13] Pieter Abbeel and Andrew Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *Proc. of ICML*, 2004.
[14] Andrew Y. Ng and Stuart J. Russell. Algorithms for inverse reinforcement learning. In *Proc. of ICML*, pp. 663–670, 2000.
[15] Radford M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, Vol. 9, No. 2, pp. 249–265, 2000.
[16] Jaedeug Choi and Kee-Eung Kim. Nonparametric Bayesian inverse reinforcement learning for multiple reward functions. In *Proc. of NIPS*, pp. 314–322, 2012.
[17] Christian Ruhhammer, Nils Hirsenkorn, Felix Klanner, and Christoph Stiller. Crowdsourced intersection parameters : A generic approach for extraction and confidence estimation. In *Proc. of IV*, pp. 581–587, 2014.
[18] Google maps. https://www.google.com/maps/.
[19] Stefan Atev, et al. Learning traffic patterns at intersections by spectral clustering of motion trajectories. In *Proc. of IROS*, pp. 4851–4856, 2006.