PREDICTION OF DRIVING ACTIONS USING DRIVING SIGNALS

Toshihiko Itoh†, Shinya Yamada†, Kazumasa Yamamoto‡ and Kenji Araki†

†Department of Information Science and Technology
Hokkaido University, Sapporo, Japan
‡Department of Information and Computer Sciences
Toyohashi University of Technology, Toyohashi, Japan
t-itoh@media.eng.hokudai.ac.jp, yamaya@media.eng.hokudai.ac.jp,
kyama@slp.ics.tut.ac.jp, araki@media.eng.hokudai.ac.jp

ABSTRACT

A spoken dialogue system for a car-navigation system has a possibility that more natural and smoother communication cause some problems concerning safety. One of some problems is a distraction. In this problem, it is pointed out that the machine operation and the voice conversation influence the driving operation. Recently, it is also pointed out that even the use of a simple speech interface may affect the driving operation. So we consider that if a spoken dialogue system can understand the driver's situations and the dialogue rhythm of the system can be changed according to the driver's situation, a safe spoken dialogue system can be achieved for a car-navigation system. For that, the system needs to predict and recognize driver's actions from environmental information such as driving signals.

In this paper, we report the result of an experiment of driver's action prediction. The driver's action prediction system uses HMM-based pattern recognition only on driving signals without position information. As a result, the best driving action prediction accuracy was 0.632.

1. INTRODUCTION

Conventional car navigation systems usually have a remote controller and a touch panel as an input and use a display output as an output. Recently, for safety reason, the speech interface is watched with keen interest, and the car navigation systems with the operation by speech commands have increased. We developed a spoken dialogue interface for a car navigation system aiming at a system that everyone can more easily use. The spoken dialogue interface can respond with dialogue rhythm of humans.

But in a spoken dialogue interface for car-navigation system, there is a possibility that more natural and smoother communication cause some problems concerning safety. One of some problems is a distraction. In this problem, it is pointed out that the machine operation and the voice conversation influence the driving operation [1, 2]. The most famous distraction problem is the influence on driving operation by speech conversation using cellular phone while driving. Recently, it is pointed out that there is a possibility that even the use of a simple speech interfaces affect the driving operation [3].

Temporal restrictions which achieve a rhythmical communication in our system maybe increase the danger of the distraction. However, we also think that the distraction by communication was caused because the dialogue partner does not understand / share driver's situation. So we consider that if a spoken dialogue system can understand the driver's situations and the dialogue rhythm of the system can be changed according to the driver's situation, a safe spoken dialogue interface can be achieved for a car-navigation system. For that, the system needs to predict and recognize driver's actions using driving signals and other information.

Pentland et al. used dynamic Markov models to recognize human behaviors from sensory data and to predict human behaviors over a few second time [4]. Oliver et al. trained Dynamical Graphical models (HMMs and CHMMs) using the experimental driving data to create models of seven different driver maneuvers: passing, changing lanes right and left, turning right and left, starting, and stopping [5].

Our research aims at the construction of a driver's action prediction system which uses simple and fast algorithm and simple data (signals) that can be acquired from usual cars, even if some accuracy is low. (This point is different from other researches.)

In this paper, we reported the result of an experiment on driver's action prediction. Our driver's action prediction system uses HMM-based pattern recognition only on driving signals without position information.

2. DRIVING SIGNALS

We used driving signals included in the CIAIR-HCC database [6] (the database of drivers' commands to in-vehicle navigators) as experimental driving signals.

2.1. The kind of driving signals

Observable driving signals can be categorized into three groups [7]:

1. Driving behavioral signals
   (e.g., gas pedal pressure, brake pedal pressure, and steering angle)
2. Vehicle status signals
   (e.g., velocity, acceleration, and engine speed)
3. Vehicle position signals

(e.g., following distance, relative lane position, and yaw angle).

Among these signals, we focused here on the driving behavioral signals and vehicle status signals that can be acquired from usual cars. We think that a vehicle position signals are very effective to predict driver's driving action, but the integration of such signals is a future work.

2.2. Database

Driving behavioral signals were collected using a data collection vehicle (Toyota Regius), which has been specially designed for data collection in the Center for Integrated Acoustic Information Research (CIAIR) project. Detailed information on this corpus can be found in [8]. Each driver drove the car on a city road, and five-channel driving signals as well as 16-channel speech signals, three-channel video signals, and GPS were recorded. The driving signals included force on gas and brake pedals, engine speed, car velocity, and steering angle shown in Table.1. These signals were sampled at 1 kHz, and 1 sample is expressed by 16 bits but quantizing level is 15 bits which use data range from 0 to 32767.

Table 1. *The data of driving behavior signal*

The driving signals	The range
Gas pedal pressure	$0 \sim 50kgf/cm^2$
Brake pedal pressure	$0 \sim 50kgf/cm^2$
Steering angle	$-1800^\circ \sim 1800^\circ$
Engine speed	$0 \sim 8000rpm$
Car velocity	$0 \sim 120km/h$

3. PREDICTION OF DRIVING ACTION

3.1. The kinds of driving actions

The goal of this research is the prediction of users' situation, to make users talk safely. For example, when a high-risk driving action is predicated to occur in a few seconds, system stops talking with driver. Therefore it is more important that system can predict whether the high-risk driving action occurs after a few time than system can recognize a present driving action.

So we think that there are not a lot of kinds of the driving action that should be predicted. Thus, The driving actions which should be predicated are follow:

1. straight ahead
2. stopping
3. turning right
4. turning left
5. changing lanes right
6. changing lanes left
7. Obstacle avoidance

The image of the driving action prediction is shown in Fig.1. The driving signals are continuously input to the system and the system predicts user's driving action after a few times.

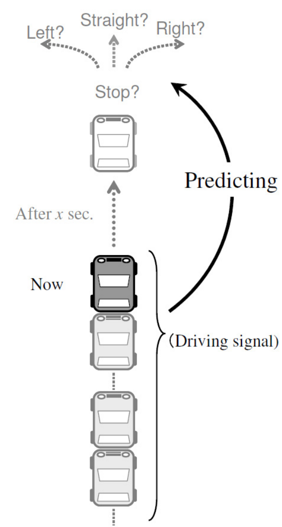

Fig. 1. *The image of the driving action prediction*

3.2. The methodology of the driving action prediction

We used an HMM-based statistical pattern recognition method for the driving action prediction. HMM-based method is widely used for the recognition of time sequence like speech, and our driving action prediction can be seen as recognition of time sequence patterns appearing before the driving actions. To be accurate, the preparation action occurring before the actual driving action we want to predict is recognized, and in consequence driving action is predicted.

Therefore we should model the preparation action preceding to each driving action. The driving signals are time series data, and HMM is well known to have strong ability to model such time sequence. In speech recognition, an input is usually a sequence of an n-dimensional vector data derived a sequence of speech frames. In driving action prediction, we used the 10-dimensional vector data in which 5 driving signal (gas pedal pressure, brake pedal pressure, steering angle, engine speed, and car velocity) and 5 regressive coefficients of them are included.

The driving action prediction should be done frame by frame as shown in Fig.2. This is a kind of detection task as so-called ``word spotting" task in the speech recognition field.

But this is the first trial of the driving action prediction and we first assure that our strategy is effective or not, we extracted the intervals preceding to some driving actions manually and conducted the experiment to classify the intervals.

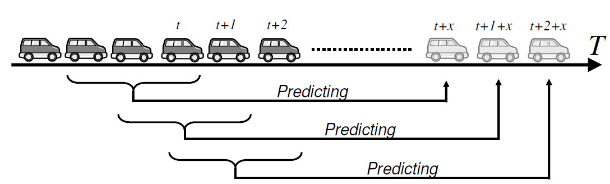

Fig. 2. *The our driving action prediciton*

4. THE DATA FOR PREDICTION EXPERIMENTS

We describe the method of making the experimental data used for the prediction experiment.

4.1. Driving action labeling

To predict the user's action described in Section 3.1, we detected the beginning and ending of driving actions manually seeing video and driving signals, and labeled ``start" and ``end" of each driving actions to the actual driving signal using an authoring tool.

Detailed features were also labeled for the research in the future. The kinds of the labels are shown in Table.2. A beginning of ``Stopping" is labeled for a moment at which a car completely stops and an ending of ``Stopping" is a moment at which a car begins to move. In addition, the kind of the label is classified by the stopping purpose and stopping state. A beginning of ``turning right", ``turning left", ``changing lanes right", ``changing lanes left" and ``Obstacle avoidance" is a moment when a steering wheel is operated and an ending is a moment at which a car begins to go straight. For these actions, the kind of the label is also classified by the purpose and state. In addition, we labeled ``passing" at a moment when a car avoided an obstacle (when a center of the obstacle is corresponding to a center of the car) for the ``Obstacle avoidance".

Table 2. *The kinds of driving action's labels*

Action	Detail	label
stopping	start for straight	Ss
	end for straight	Se
	start for turn	S[L,R]s
	end for turn	S[L,R]e
	start for turning	S[L,R]Ds
	end for turning	S[L,R]De
turning right	start	Rs
	end	Re
	start after Stop	RSs
	end after stop	RSe
turning left	start	Ls
	end	Le
	start after Stop	LSs
	end after stop	LSe
changing lanes right	start	CRs
	end	CRe
changing lanes left	start	CLs
	end	CLe
Obstacle avoidance	start	As
	passing	A
	end	Ae

4.2. Training data and test data

We made training data and test data for the driving action prediction experiments by the following process (Fig.3). Our task is to predict the driving action occurring at time X and the prediction should be done Y seconds before the occurrence. Now we assume that the system can use the past Z -second signals to predict.

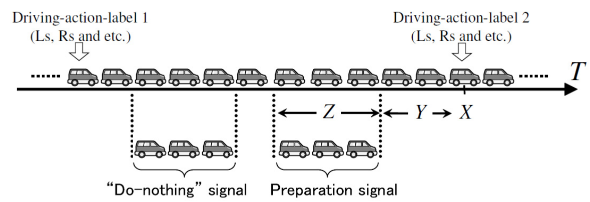

Fig. 3. *The extraction of experimental driving signals*

Table 3. *The number of extracted experimental driving action data*

Driving Action	# Total data
straight ahead	1570
stopping	764
turning right	66
turning left	184
changing lanes	141
Obstacle avoidance	1426

So we extracted the driving signals from $(X-Y-Z)$ to $(X-Y)$. We also extracted the regions without any special driving actions following them, as the data from which the system should predict the ``straight ahead" action.

Thus, we obtained such ``preparation signals" for driving actions from the driving data of 110 subjects. Preparation signals from 70 subjects are used as training data and those from 40 are used as test data. Details of the number of the data are shown in Table.3.

This paper reports only the results of a preliminary experiment in which the action happening just after the prediction (In other word, the constant Y is 0).

5. THE RESULTS OF THE PREDICTION EXPERIMENTS

HTK [11] was used for all HMM training and the driving action prediction experiment.

5.1. Prediction performance comparison by differences of the input duration of driving signals (experiment 1)

First, we investigated the change in the prediction performance by the difference of the input duration of driving signals. Table.4 shows the experimental conditions. We prepared 1, 5 and 10 seconds' driving signals for the prediction. The signals from 20 subjects are used as training data and those from other 40 are used as test data.

Table.5 shows the accuracy of the experiment. As a tendency, the duration of the input signals is longer, the prediction accuracy is better. The performance with 5 seconds' input and 10 seconds' input was almost the same. And when the duration of the input was 10 seconds and the number of mixture was 32, the accuracy of the driving action prediction was decreased.

Table 4. *The condition of the experiment 1*

The experimental condition	
HMM	Ergodic HMM
The number of state	4
The number of mixture	2,4,8,16,32
Feature parameters	5 driving signals + 5 Δdriving signals
Duration of the input	1, 5, 10 sec
# Training data	20 subjects
# Test data	other 40 subjects

Table 5. *The accuracy of the experiment 1*

# mixture	the duration of input		
	1 sec	5 sec	10 sec
2	0.470	0.502	0.522
4	0.488	0.554	0.521
8	0.546	0.587	0.544
16	0.559	0.554	0.579
32	0.550	0.557	0.527

Table 6. *The confusion matrix of the experiment 1 (the mixture was 16 and the duration of the input was 10 seconds.)*

	OA	CL	TL	SA	TR	SP	total	precision	F-value
OA	196	15	19	117	2	1	351	0.558	0.577
CL	25	12	2	31	0	0	70	0.171	0.198
TL	12	8	17	79	2	2	123	0.138	0.180
SA	76	12	16	298	8	2	422	0.706	0.592
TR	8	1	12	21	9	8	62	0.145	0.207
SP	11	3	0	39	4	246	286	0.801	0.840
total	328	51	66	585	25	259	1314		
recall	0.598	0.235	0.258	0.509	0.360	0.950		0.579	

We think that training data was not enough. Further, when the duration of the input was 10 seconds, the training time of HMMs increases considerably. So we think that it is enough in the duration of input signals is about 5 second.

Table.6 shows the confusion matrix of the prediction result with the highest accuracy. In the table, the "OA" is the ``Obstacle avoidance", the ``CL" is the ``changing lanes (left or right)", the ``TL" is the ``turning left", the ``TR" is the ``turning right" and the ``SP" is the ``stopping". Subheadings should appear in lower case (initial word capitalized) in boldface. They should start at the left margin on a separate line.

5.2. Prediction performance comparison by differences of the number of training data (experiment 2)

We investigated the change in the prediction performance by the difference of the training data.

Table.7 shows the experimental conditions and Table.8 shows the accuracy of the experiment. The more training data we used, the higher accuracy we obtained. The highest accuracy was achieved by the model with 32 mixture trained using 10-seconds duration data of all the subjects.

Table.9 shows the confusion matrix of the prediction result with the highest accuracy in this experiment.

Table 7. *The condition of the experiment 2*

The experimental condition	
HMM	Ergodic HMM
The number of state	4
The number of mixture	2,4,8,16,32
Feature parameters	5 driving signals + 5 Δdriving signals
Duration of the input	5 sec, 10 sec
# Training data	20, 40 , 70 subjects
# Test data	other 40 subjects

Table 8. *The accuracy of the experiment 2*

# mixture	The duration of input				
	5 seconds		10 seconds		
	The amount of training data		The amount of training data		
	20 subjects	70subjects	20 subjects	40 subjects	70subjects
2	0.502	0.536	0.522	0.550	0.567
4	0.554	0.480	0.521	0.522	0.559
8	0.587	0.501	0.544	0.557	0.549
16	0.554	0.517	0.579	0.590	0.573
32	0.557	0.487	0.527	0.550	0.632

Table 9. *The confusion matrix of the experiment 2 (the mixture is 32, the duration is 10 seconds, the training data is 70 subjects)*

	OA	CL	TL	SA	TR	SP	total	precision	F-value
OA	227	13	9	128	2	1	380	0.597	0.641
CL	35	30	4	53	0	0	122	0.246	0.347
TL	9	3	43	73	1	2	131	0.328	0.437
SA	31	2	5	267	3	2	310	0.861	0.597
TR	14	1	5	38	17	8	83	0.205	0.315
SP	12	2	0	26	2	246	288	0.854	0.899
total	328	51	66	585	25	259	1314		
recall	0.692	0.588	0.652	0.456	0.680	0.950		0.632	

5.3. Prediction performance that considers individuality of driving (experiment 3)

So far we used the training data not including the persons who are included in the test data. But it should be pointed out that a way of the driving is greatly different depending on the person. So we conducted the experiment that the signals from 2/3 of each person are used as training data and the rest as test data.

Table.10 shows the experimental conditions and Table.11 shows the accuracy of the experiment. When compared with the results of the models trained using 5-seconds duration data of 70 subjects in Table.8, the accuracy has improved though almost same amount of data is used. This reveals that it is important to consider the individuality of driving action.

Table.12 shows the confusion matrix of the prediction result with the highest accuracy in this experiment.

5.4. Prediction performance using only one signal of 5 driving signals (experiment 4)

Next, we investigate the contribution level to the prediction of an individual driving signal. So we conducted the experiment in which only one signal of the driving signals (gas pedal pressure, brake pedal pressure, engine speed, car velocity and steering angle) was used.

Table 10. *The condition of the experiment 3*

The experimental condition	
HMM	Ergodic HMM
The number of state	4
The number of mixture	2,4,8,16,32
Feature parameters	5 driving signals + 5 Δdriving signals
Duration of the input	5 sec
# Training data	2/3 of 110 subjects
# Test data	1/3 of 110 subjects

Table 11. *The accuracy of the experiment 3*

# mixture	accuracy
2	0.522
4	0.515
8	0.505
16	0.543
32	0.529

Table 12. *The confusion matrix of the experiment 3 (the mixture is 16.)*

	OA	CL	TL	SA	TR	SP	total	precision	F-value
OA	180	12	9	134	2	2	339	0.531	0.570
CL	25	17	8	45	0	0	95	0.179	0.238
TL	35	11	39	89	9	1	184	0.212	0.317
SA	39	7	1	168	2	7	224	0.750	0.449
TR	10	1	5	35	7	2	60	0.117	0.169
SP	4	0	0	53	3	245	305	0.803	0.872
total	293	48	62	524	23	257	1207		
recall	0.614	0.354	0.629	0.321	0.304	0.953		0.543	

Table.13 shows the experimental conditions and Table.14 shows the result of the experiment in accuracy. The contribution level to the prediction was greatly different depending on the signal. For the driving action prediction, we noticed that steering angle, car velocity and engine speed are greatly useful pieces of information, and gas pedal pressure and brake pedal pressure is not. Certainly, these signals are greatly changed by the driving situation than the driving action. These signals may negatively affect the driving action prediction.

Table.18 shows the confusion matrix of the prediction result with the highest accuracy in this experiment.

5.5. Prediction performance using useful 3 signals out of 5 (experiment 5)

The gas pedal pressure and the brake pedal pressure are easy to change greatly in any driving action. These signals may negatively affect the driving action prediction.

So, we used only steering angle, car velocity and engine speed which are greatly useful pieces of information.

Table.16 shows the experimental conditions and Table.17 shows the accuracy of the experiment. A prediction performance is the highest when duration of the input signals is 5 seconds. From this result, because the fluctuation (change) of each driving action is large by an influence of individuality and situation, the models of HMM can not model each driving action using many input signals.

Table 13. *The condition of the experiment 4*

The experimental condition	
HMM	Ergodic HMM
The number of state	4
The number of mixture	2,4,8,16,32
Feature parameters	individual driving signal
Duration of the input	5 sec
# Training data	40 subject
# Test data	same 40 subject

Table 14. *The accuracy of the experiment 4*

# mixture	the driving signal				
	gas	brake	engine	velocity	steering
2	0.203	0.209	0.352	0.436	0.238
4	0.225	0.303	0.370	0.431	0.341
8	0.243	0.278	0.377	0.387	0.333
16	0.348	0.205	0.356	0.437	0.349
32	0.201	0.206	0.362	0.438	0.424

Table 15. *The confusion matrix of the experiment 4 (the mixture is 32 and input signal is car velocity.)*

	OA	CL	TL	SA	TR	SP	total	precision	F-value
OA	174	18	27	163	3	5	390	0.446	0.506
CL	29	20	3	52	0	11	115	0.174	0.227
TL	31	7	20	63	8	1	130	0.154	0.207
SA	62	14	10	249	7	153	495	0.503	0.481
TR	1	2	1	10	5	11	30	0.167	0.189
SP	1	0	2	3	0	76	82	0.927	0.448
total	298	61	63	540	23	257	1242		
recall	0.584	0.328	0.317	0.461	0.217	0.296		0.438	

Table.18 shows the confusion matrix of the prediction result with the highest accuracy in this experiment.

5.5. Prediction performance by the detailed classification of the driving actions (experiment 6)

From the results above, a lot of variations of the patterns were included in a kind of action. So we divided the each driving action into some patterns by similarity of driving signals using clustering tool (the ``straight ahead" was divided into 4 sub-classes, the ``stopping" was 3, the ``turning right" was 2, the ``turning left" was 4, the `changing lanes right and left" was 4, and the ``Obstacle avoidance" was 4.). An HMM of each sub-class was trained using the only driving signals included in the sub-class. The class one of whose sub-class has the highest likelihood is selected as the predicted action.

Table.19 shows the experimental conditions and Table.20 shows the accuracy of the experiment. The performance was the highest when duration of the input signals is 5 seconds.

Table.21 shows the confusion matrix of the prediction result with the highest accuracy in this experiment.

Table 16. *The condition of the experiment 5*

The experimental condition	
HMM	Ergodic HMM
The number of state	4
The number of mixture	2,4,8,16,32,64
Feature parameters	3 driving signals (engine, velocity, steering)
Duration of the input	5 sec
# Training data	40 subjects
# Test data	other 40 subjects

Table 17. *The accuracy of the experiment 5*

# mixture	accuracy
2	0.475
4	0.525
8	0.537
16	0.571
32	0.584
64	0.594

Table 18. *The confusion matrix of the experiment 5 (the mixture is 64.)*

	OA	CL	TL	SA	TR	SP	total	precision	F-value
OA	244	29	16	181	3	3	476	0.513	0.600
CL	9	7	3	22	0	0	41	0.171	0.151
TL	19	1	39	21	4	0	84	0.464	0.517
SA	60	13	9	264	11	26	383	0.689	0.545
TR	3	0	0	9	4	4	20	0.200	0.178
SP	2	2	0	88	3	233	328	0.710	0.785
total	337	52	67	585	25	266	1332		
recall	0.724	0.135	0.582	0.451	0.160	0.876		0.594	

Table 19. *The condition of the experiment 6*

The experimental condition	
HMM	Ergodic HMM
The number of state	4
The number of mixture	2,4,8,16,32,64
Feature parameters	5 driving signals + 5 Δdriving signals
Duration of the input	5 sec
# Training data	70 subjects
# Test data	other 40 subjects

Table 20. *The accuracy of the experiment 6*

# mixture	accuracy
2	0.487
4	0.533
8	0.561
16	0.593
32	0.615
64	0.612

Table 21. *The confusion matrix of the experiment 6 (the mixture i 32.)*

	OA	CL	TL	SA	TR	SP	total	precision	F-value
OA	266	31	16	197	4	1	515	0.517	0.627
CL	13	6	1	19	0	0	39	0.154	0.133
TL	14	2	32	54	3	1	106	0.302	0.370
SA	31	8	13	255	9	12	328	0.777	0.559
TR	3	0	3	2	5	0	13	0.385	0.263
SP	6	4	2	58	4	251	335	0.772	0.851
total	333	51	67	585	25	265	1326		
recall	0.799	0.118	0.478	0.436	0.200	0.947		0.615	

6. CONCLUSION

In this paper, we reported the result of an experiment on driver's action prediction. The driver's action prediction system uses HMM-based pattern recognition only on driving signals without position information. As a result, the best driving action prediction accuracy was 0.632. Considering that the driving action prediction system used only a very simple framework and input signals, the result was very promising.

We expect that a vehicle position signals is very effective to predict driver's driving action. The information of the position signals may play the role as a grammatical constraint in the speech recognition field. And we will try other methods (algorithms) to predict driving actions.

7. REFERENCES

[1] J. D. Lee, T. L. Brown B. Caven, S. Haake, K. Schmidt, “Does a speech-based interface for an in-vehicle computer distract drivers?”, Proc. World Congress on Intelligent Transport System (2000).

[2] http://www-nrd.nhtsa.dot.gov/departments/nrd-13/DriverDistraction.html

[3] http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Welcome.htm

[4] A. Pentland and A. Liu, “Modeling and prediction of human behavior”, Neural Computation, vol. 11, pp. 229.242, 1999.

[5] N. Oliver and A.P. Pentland, “Driver behavior recognition and prediction in a SmartCar”, Proc. SPIE Aerosense 2000, Enhanced and Synthetic Vision, Apr. 2000.

[6] http://db.ciair.coe.nagoya-u.ac.jp/eng/index.html

[7] Y. Nishiwaki, K. Ozawa, T. Wakita, C. Miyajima, K. Itou, and K. Takeda, ``Driver identification based on spectral analysis of driving behavioral signals", Proc. Biennial on DSP for in-Vehicle and Mobile Systems, 2005.

[8] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, “Multimedia data collection of in-car speech communication”, Proc. EUROSPEECH 2001, pp. 2027.2030, Sept. 2001.

[9] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev and P. Woodland, ``The HTK book -version2.2", Entropic, 1999.