Acquisition of Probabilistic Behavior Decision Model based on the Interactive Teaching Method

Tetsunari INAMURA  Masayuki INABA  Hirochika INOUE

Inoue-Inaba Lab., Dept. of Mechano-Informatics, Faculty of Engineering, The University of Tokyo. 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN

Abstract

In this paper, we propose a novel method for mobile robots to acquire new autonomous behaviors gradually based on interaction between human and robots. In this method, behavior decision models are constructed using statistical process for experiences of interaction and teaching, and the robot expresses the robot-on decision using stochastic reasoning. The robot is not only decides behavior using the robot-inference, but also makes suggestions and questions for the user using the robot-inference. Consequently, the robot-on inference enables the behavior acquisition to be more effective. We refer to this kind of method as “Interactive Teaching” method. We investigate the feasibility of this method for obstacle avoidance tasks with mobile robots. Through experiments at both virtual and real mobile robot, we have confirmed that the mobile robot acquires robust behavior decision models against changes of environment and uncertainties of sensors, through only several teaching.

1 Introduction

We have studied on personal robots which can act in daily life environments to support human. In such a situation, well-found infrastructure cannot be expected, the robot has no knowledge at the beginning of use, and a user gives the robot many requests with vague expression. The robot must act robustly in such unfamiliar environments with uncertain information. However a developer cannot know details of the environments in advance, therefore the developer cannot embed certain knowledge in the robot. To overcome such difficulties, adaptation is required. For the adaptation, it is to be desirable that personal robots make the best use of interaction between human and themselves, because of its concept, collaboration with human. Namely, the function of incremental acquisition of autonomous behavior while gaining experience through interaction, is important to personal robots.

Torrance[1] developed an interactive mobile robot which accepts some requests from users via natural language, and behaves following the instruction. However this system only changes behaviors using an instruction-behavior relation table embedded beforehand. Thus it is difficult for the user to alter the relation table dynamically. Shibata et al.[2] developed a real-time interactive system on a mobile robot. In this system, a user board teaches the robot how to move in an unfamiliar corridor environment. This system is more easier to alter the behavior because several basic behaviors have been embedded and the user determines the combination of these behaviors. Nevertheless the system can not cope with changes of environments. If the change occurs, the robot must be taught all over again. This problem is caused by fixed expression of relation between visual information and behaviors, as same as the teaching by showing which is Kuniyoshi’s work[3].

Jijo-2 project[4] has proposed a framework in which a mobile robot acquires probabilistic environment model through real-world interaction. However the framework has not touched on the acquisition of behavior model.

Therefore the aim of this research is to realize an automatic acquisition of behavior model through interaction between human and robots. Features of the model are as following:

- The model is not embedded in the robot beforehand, but easy to be understood.
- The model is not stiff but flexible for user’s dynamical teaching,
- Learning based on not one-way teaching, but interaction between human and robot.

We refer to this kind of interaction method as Interactive Teaching.

In section 2, we mention a framework for automatic behavior model acquisition through observation of user’s teaching operation. In section 3, we touch on an interaction method between robots and human during the teaching. In section 4, we explain an application of the Interactive Teaching to obstacle avoidance tasks. We also show an actual experiment system on a virtual mobile robot environment, where a user can interact with the robot using interface devices such as lever joystick, microphone, speaker, and so on. In section 5, we show the validity of our method through various kinds of experiments. Section 6 is the conclusion.
2 Stochastic Behavior Decision
Model based on User’s Teaching

In this section, we explain the flexible behavior decision model. The defect of the stiff model is weakness for the changes of environment, as mentioned above. It also has an aspect that recycling old model against new environment is difficult. A user’s operation also might be imperfect, thus the model formation follows user’s mistake.

It is to be desired that the stochastic expression is used without the fixed expression. The reasons for adoption of stochastic expression are:

1. On-line learning is available using simple cumulative frequency calculation from interaction experience.

2. The sureness of decision can be expressed as probability value. It also be realized that the robot controls interaction using the sureness.

3. Stochastic process can cancel the influence of noise such as mistake of the user.

We have adopted Bayesian Network[5][6] in order to build such a stochastic model. Bayesian Network is a reasoning model in which the relation between cause and effect of plural phenomena are expressed as probability. It is used in various fields such as map acquisition for mobile robots [7][8], diagnosis for help system on personal computers[9], dialogue management[9], and so on. In Bayesian Network, phenomena are expressed by nodes and relations between these phenomena are expressed with arcs as shown in Fig. 1. Each node has a random variable which is the probability of the phenomenon. Relations between cause and effect corresponding with each arc are expressed by conditional probabilities. Nodes are classified roughly into two kinds: evidence nodes which correspond with the evidence phenomenon to be used in reasoning, and hypothetical nodes which correspond with the object phenomenon to be deduced. In the case of robot applications, an evidence node indicates a sensor of the robot, and a hypothetical node indicates behavior. The behavior node expresses a behavior which will be executed through reasoning.

Let $E = \{e_1, e_2, \ldots, e_n\}$ be the random variable of an evidence node, and let $H = \{h_1, h_2, \ldots, h_m\}$ be the random variable of a hypothetical node. A causal relationship between evidence node $E$ and hypothetical node $H$ are described by following matrix:

$$CPT_{E|H} \overset{\text{def}}{=} P(E|H)$$

$$= \begin{pmatrix}
P(e_1|h_1) & P(e_2|h_1) & \ldots & P(e_n|h_1) \\
P(e_1|h_2) & P(e_2|h_2) & \ldots & P(e_n|h_2) \\
\vdots & \vdots & \ddots & \vdots \\
P(e_1|h_m) & P(e_2|h_m) & \ldots & P(e_n|h_m)
\end{pmatrix} \quad (1)$$

This matrix is called as Conditional Probabilities Table (CPT). Conversely when the robot acts autonomously, the random variable of a hypothetical node is calculated after the observation of evidence nodes. The calculation is expressed as follows:

$$BEL(H) \overset{\text{def}}{=} \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{CPT_{E|H} \cdot P(H)}{P(E)} \quad (2)$$

When this model is applied to robots, the evidence node $E$ is regarded as sensor node $S$, and the hypothesis node $H$ is regarded as behavior node $B$. The robot acquires prior probability $P(B)$ and causal relation $CPT_{S|B}$ through interaction experience. When the robot moves autonomously, the posterior probability $P(B|S)$ is calculated by observation of sensor node $P(S)$.

The result vector $BEL(H)$ is called the degree of sureness whose constituents indicate the plausibility for each status of the hypothetical node, namely the behavior option. In case that the degree of sureness becomes $BEL(B) = (0.01, 0.97, 0.02)$, it indicates that the robot is sure of executing the second behavior. Conversely when the degree of sureness $BEL(B) = (0.0, 0.5, 0.5)$ is deduced, it indicates that the second and the third behavior are suitable with fifty-fifty sureness. In such case, the robot can make a question to the user, “Please teach me which is the better selection?” It is also possible that the robot points out the user’s mistake, when the degree of sureness is extremely high, and the user’s instruction differs from decision of the robot.

3 Incremental Behavior Acquisition
based on Interactive Teaching

In this section, we explain how to acquire CPT and prior probabilities effectively, using interaction between human and robots. Let $N$ be the number of pieces of observed experience, and $N_j$ be the number of human teaching behavior $h_j$, which occurred among $N$ times experiences. Also, let $n_{ij}$ be the number of sensor information $s_i$ was observed in the $N_j$ times. In this case, the most simple calculation is
### Table 1: Interaction processes for each situation

<table>
<thead>
<tr>
<th>Degree of Ssureness [%]</th>
<th>The differences between user's teaching and reasoning results of robot</th>
<th>Case in human doesn't teach</th>
</tr>
</thead>
<tbody>
<tr>
<td>80~100</td>
<td>Obey user's operations</td>
<td>Execute reasoning result</td>
</tr>
<tr>
<td>Autonomous Level 4</td>
<td>Do not store</td>
<td>Do not store</td>
</tr>
<tr>
<td>60~79</td>
<td>Obey user's operations</td>
<td>Execute reasoning results</td>
</tr>
<tr>
<td>Autonomous Level 3</td>
<td>Storage with weight 1</td>
<td>while showing deduced choices</td>
</tr>
<tr>
<td>60~64</td>
<td>Obey user's operations</td>
<td>Execute reasoning results</td>
</tr>
<tr>
<td>Autonomous Level 2</td>
<td>Storage with weight 3</td>
<td>while showing deduced choices</td>
</tr>
<tr>
<td>0~49</td>
<td>Obey user's operations</td>
<td>Wait for user's operations</td>
</tr>
<tr>
<td>Autonomous Level 1</td>
<td>Storage with weight 5</td>
<td>while showing deduced choices</td>
</tr>
</tbody>
</table>

\[
P(s_i|b_j) = P(S = s_i|B = b_j) = \frac{n_{ij}}{N_j} \quad (3)
\]

\[
P(b_j) = P(B = b_j) = \frac{N_j}{N} \quad (4)
\]

There arises a problem when this simple cumulative frequency is used. When rare behavior was taught, the prior probability \( P(B) \) is partial to the major behavior. Therefore, the sureness of decision for the rare behavior is less than plausible value. This contradiction is caused by the assumption that uniform probability distribution is used for each behavior and sensor node. It is supposed that psychological experience accumulation is expressed using not mere frequency of encounter but weighted frequency with importance of situation. In this paper, we assume that the important situation depends on following two factors: (1) Whether the user’s teaching differs from the robot’s reasoning result. (2) The sureness of behavior decision. Therefore, we introduce degree of importance for each situation as Table 1.

For example, when the robot outputs high degree of sureness and the user doesn’t teach the robot, this situation is regarded as a familiar situation and experience data is not stored. Conversely, when the degree of sureness is low and the user’s operation is differ from decision of the robot, this situation is regarded as unfamiliar situation and experience data is stored with weight 10. The latter case is more important than the former, thus it ought to be focused on for adaptation and learning.

Here, we explain an algorithm of CPT calculation with weight shown in Table 1. Let \( N_j^{(b)} \) be the number of situation \( C_b \) while the user operated behavior \( b_j \) for \( N_j \) times. Similarly, let \( n_{ij}^{(b)} \) be the number of situation \( C_b \) while both behavior \( b_j \) and sensor \( s_i \) were observed at same time. In case that the weight of situation \( C_b \) is \( w_b \), variables are expressed with \( t \) as following.

\[
N_j = \sum_k w_k N_j^{(b)} \quad (5)
\]

\[
in'_{ij} = \sum_k w_k n_{ij}^{(b)} \quad (6)
\]

\[
P'(s_i|b_j) = \frac{n'_{ij}}{N_j'} \quad (7)
\]

\[
CPT'_{S|B} = P'(S = s|B = b) \quad (8)
\]

\[
P'(B) = P'(B = b_j) = \frac{N_j'}{\sum_j N_j'} \quad (9)
\]

The robot changes level of autonomy according to the degree of sureness and the difference between reasoning results and user’s teaching. Through questions, suggestions and presentations, shown in Table 1, the robot stores its experience with each weight, and learns the CPT matrix using Eq.(1)(8)(9). The procedure of the calculation is illustrated in Fig.2. We refer such a method that the robot learn behavior based on experience through interaction with human, as Interactive Teaching. In the Interactive Teaching, the learning phase and the autonomous running phase are not distinguished. In brief, the robot is under the semiautomated control of the user.
4 Application for Obstacle Avoidance Tasks on Mobile Robots

4.1 Model Expression via Bayesian Network

In an obstacle avoidance task for a mobile robot in this research, the teaching information given by users consists of direct instructions via joystick levers and indirect instruction via language. The user instructs the robot via language when the robot makes a question, and make a suggestion with enumerating some behavior options.

A direct instruction via joystick lever corresponds with the behavior node. Considering obstacle avoidance tasks, we have adopted distance sensors for the sensor nodes. A Bayesian Network used in the avoidance application is shown in Fig. 3. The robot measures distance along with horizontal line as shown in Fig.4. After the measurement, the result is expressed with eight values which correspond with each sensor node. Each distance sensor value is an integer. Zero means closest and 255 means farthest distance. These distance values are translated to state values because the sensor information is expressed as random variable at each sensor node.

We used a random variable $S_i = \{Near, Far\}$ for each sensor node. For the behavior node which expresses user’s operation, a random variable $B = \{Forward, Stop, Left, Right\}$ is used. After the observation of these sensor node, namely the acquisition of CPT matrix, reasoning is executed for the behavior node. The reasoning uses following formalization

$$BEL(B) = \alpha \prod_i \lambda_i(B) \pi(B)$$  \hspace{1cm} (10)

where $\alpha$ is a coefficient for normalization, $\pi(B)$ is prior probability of the behavior node, and $\lambda_i(B)$ is the likelihood vector which is calculated by

$$\lambda_i(B) = CPT_{S_i|B} \lambda(S_i)$$  \hspace{1cm} (11)

where $\lambda(S_i)$ is sensor input of $i$-th sensor node. See the literature[3] for detail calculation processes.

4.2 Experiment System using Virtual Mobile Robot Environment

In order to carry out experiments of the Interactive Teaching for obstacle avoidance tasks, we have developed a virtual environment for mobile robots and a user interface terminal (Fig.5).

The virtual environment is prepared using OpenGL library on IBM PC/AT clone. The distance sensors are emulated using Z buffer of the 3D model. The user interface terminal has a joystick, a headphone with a microphone for speech interaction, a keyboard and so on. User’s utterance are transformed into symbols through speech recognition subsystem. After the recognition, symbols are sent to interaction manager. The interaction manager controls the mobile robot, following the user’s requests such as “go forward” and “turn right”. When the robot tries to make questions, suggestions and presentations, speech synthesizer subsystem generates speech utterance.

5 Experiments of the Interactive Teaching

We have practiced the Interactive Teaching experiment in a virtual corridor shown in Fig. 6. The user operates the mobile robot from the start point to the destination point which is the end of the corridor. In the first step, the robot does not know how to avoid obstacles, the size of its body, and where should it stop. Thus, the user operates the robot along with entire path at the first time.

Fig. 7 shows a progress graph of learning. The horizontal axis is the number of trial running. In this case, “one time” is a running from the start to the end of the corridor. The vertical axis means the de-
gure of sureness. As the graph shows, the robot can select suitable behavior for all scene. The learning converged after three times running. Moreover the user hardly instructed any operations in the second trial. In this time, the robot almost ran by itself, and the user corrected only a few wrong behavior. These results indicate the validity of this method.

5.1 Coping with Dynamic Change of Environment

This method uses stochastic expression for the behavior decision model, thus it has robustness property for dynamic change of environment. Even though the robot learned in an environment in which all obstacles are fixed, the robot can flexibly cope with modified environments.

We have made some trials of running in environment B shown in Fig.8, after learning in environment A shown in Fig.6. The solid line means the result when the robot used weighting storage strategy expressed by Eq.(5)~(9). It is obvious that the robot can arrive at the destination point without hindrance. The broken line shows the result when the robot used normal storage strategy expressed by Eq.(3)(4).

When the weighting storage strategy was used, the robot could avoid obstacles normally. However the normal storage strategy was used, the robot collided without appropriate avoidance. At this time, the prior probability has been changed from $P(B) = (0.81, 0.037, 0.041, 0.064)$ to $P(B) = (0.73, 0.002, 0.064, 0.106)$. It signifies that the robot could not avoid obstacles because the degree of sureness for “turn right” was smaller than the one at weighting storage strategy. Therefore we see that the interactive weighting storage strategy is better than normal one.

We have also carried out two experiment that the robot starts with different location and direction. These condition is differ from the one where the robot went through at the learning stage. The result of this experiments is shown in Fig.9. In either case, the robot is able to avoid obstacles.

5.2 Coping with the Uncertain Sensor Information

This method can select suitable behavior robustly not only when the environment changes but also when the sensor information is ambiguous. Here, we shall argue a case where sensors are broken.

In the framework of Bayesian Network, reasoning can be executed while some nodes are unobserved. Thus the robot can deduce without input from broken sensor nodes. Results of this imperfect reasoning are shown in Fig.10. The line (a) indicates the track where second sensor from the left and second sensor from the right are broken. The line (b) indicates where second and third sensor from the left and third sensor from the right are broken. The line (c) indicates where four sensors in the middle are broken. As these results indicate, the robot could not arrive at the destination in cases (b) and (c), however could keep the avoidance ability in case where 25% of all sensors are broken.

5.3 Experiment in Real Environment

Fig.11 indicates the progress graph of learning on a real mobile robot[10]. The meaning of each axis
are same as described before. The convergence of the learning becomes dull, because of the noise in the real environment. However the suitable behavior was selected after four teaching operations. This result indicates that the Interactive Teaching method has equal validity in real environment as in virtual environment.

6 Conclusion

We have described the acquisition method of stochastic behavior decision model based on interaction between human and robot. In order to investigate the validity of this method, we have focused on obstacle avoidance tasks for mobile robots. The robot could acquire the obstacle avoidance model by means of integration of Bayesian Network framework and the Interactive Teaching. It was observed that (1) effective model acquisition is available with the help of the Interactive Teaching method, (2) robust behaviors can be generated to cope with changes of environment, and (3) robust behavior can also be generated to cope with uncertain sensor information such as broken sensor.

In the present circumstance, the reasoning process is simple, thus only sensor information at certain moments is used. Sensor information for time series should be used, however that remains to be done. An expansion method for Bayesian Network which can deal with the time series information has been proposed[11]. We suppose that this kind of expansion will be very important.

Even though in other cases from obstacle avoidance tasks for mobile robots, it is able to cope with some applications by means of making suitable structure of Bayesian Network models. A problem now arises that network structure should be described by developer beforehand. Toward that, automatic structure generation methods are being discussed. We also making a study for structure generation, and intending to announce the result in another paper.

The important factor for knowledge and behavior acquisition systems is not the perfect autonomy embedded from the beginning, but the cooperative interaction between users and the robot. We are planning to advance this research, with applying the Interactive Teaching to some applications which need cooperative interaction between human and robots.

![Image](image-url)

**Fig. 10:** Results when sensors were broken

**Fig. 11:** Result on a real mobile robot

**References**