The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

Management of Experience Data for Rapid Adaptation to New Policies based on Bayesian Significance Evaluation

Saifuddin Md. Tareeq and Tetsunari Inamura
The Graduate University for Advanced Studies, National Institute of Informatics, Japan
smtareeq@nii.ac.jp;inamura@nii.ac.jp

Abstract

This paper shows a rapid learning method of behavior policy for mobile robots teleoperated by an operator. Rapid policy adaptation cannot be achieved when data from every process cycle is used for learning because important and meaningful data are not differentiated with other data. We propose a method to solve the problem by selecting significant data for the learning based on change in degree of confidence of the behavior decision. A small change in the degree of confidence can be regarded as reflecting insignificant data for learning, so that data can be discarded. Accordingly the system can avoid having to store too much experience data and the robot can adapt rapidly to changes in the user's policy. In this paper we discuss the experimental result of an experiment in which user policy changes between 'avoid' and 'approach' on a mobile robot.

Keywords: Bayesian Network, Rapid Adaptation, Degree of Confidence

1 Introduction

One of the important abilities for personal service robots which act in real environment with human beings is to learn and acquire novel behavior strategy according to observation of users' behavior. To learn the behavior strategy, conventional methods observes sets of sensor input and command output, extracts meaningful relation between the sensor and commands using statistical methods. But the performance of the learning strongly depends on the quality of the dataset of sensor and command. When the dataset included important and meaningful experience data, the learning would be a success; however it is difficult to obtain sophisticated experience dataset for human-robot interaction in real world, because the robots basically stores the dataset in every process cycle. For example, when a user kept operating same command in same situation, the statistical learning procedure tends to output the frequent command even though the sensor is not the frequent but rare. To select the rare command for rare sensor, the system should ignore insignificant frequent dataset to avoid bad learning quality. In this paper, we propose a technique to manage experience dataset with evaluation of significance of the dataset based on a concept of change in degree of confidence for behavior decision. A small change can be regarded as an insignificant data for learning, so that data will be discarded. Accordingly, the system can avoid having to store too frequent experience data.

Conventional methods like window [1][2] based adaptation require background investigation of the domain to find a suitable window, dual model [3][4] based methods uses separate model for short term and long term learning but are unsuitable for rapid adaptation with long term model, interaction [5][6][7] based methods does not deal with adaptation with user policy but only with acquiring user policy.

Bayesian network is suitable to represent policy, because sensor and command can be represented even though the observation of the user is not well conducted and also it can output a degree of belief for behavior decision based on observation of sensor as evidence. Conventional simple belief calculation based on frequency of the dataset causes the problem that the system tends to output the most frequent command even though sensor input for rare situation is given, when the dataset observed continuously during the human-robot interaction. The problem arises because the prior probability is calculated using the numbers of observations. This factor also causes another problem that the robot cannot adapt rapidly to changeable policies of the user. We adopt Dirichlet distribution to evaluate the significance of data. The Dirichlet distribution represents not only event probability among several propositions, but also degree of confidence for the output probability just referring a set of number of observation for the propositions. The system calculates the degree of confidence before and after the current observation. The change in the two degrees of confidence can be regarded as the importance of the observation to the learning process.

2 Bayesian Network and Significance Evaluation

2.1 Bayesian network

A Bayesian network is a directed acyclic graph consists of parent nodes representing causes and child nodes representing effects as shown in Fig. 1. Each node has a propositions assigned to it which might have several values. The activation of a proposition is represented probabilistically, and as a result, each node has a stochastic variable. Specifically, sensor information in the robot, the behavior that the robot is to perform, and the content provided by a person are assigned to a node. The relationship among nodes is described using conditional probabilities with stochastic variables.

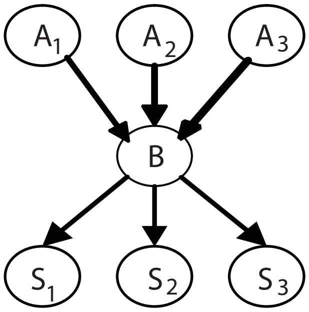

Figure 1: A Bayesian network

Let us denote a stochastic variable for a particular node B with the propositional symbols $b_1, b_2, \ldots, b_m$ by $B = \{b_1, b_2, \ldots, b_m\}$. The set of parent nodes connected from above to this node is $\mathbf{A} = \{A_1, A_2, \ldots, A_k\}$, and the space consisting of a combination of each value for the stochastic variables is $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_l$. The reasoning can be expressed

$$Bel\,(B) = \beta\lambda\,(B)\,\pi\,(B)\,, \qquad (1)$$

where $\lambda\,(B)$ represent the current strength of diagnostic support contributed by the children of B given by $\prod_i \lambda_i\,(B)$, $\pi\,(B)$ represent the current strength of the causal support contributed by the parents of B and β is the coefficient for normalization. Elements of $Bel\,(B)$ indicate the plausibility for each proposition of the behavior node. One of the advantages of Bayesian networks is that a robot can evaluate the vagueness of a behavior decision, and this leads it to ask questions and give suggestions to users [5]. For example, the robot should ask the user to confirm the behavior decision when the elements of $Bel\,(B)$ are almost equal.

2.2 Conventional Method

Suppose that B represent a behavior node and it does not have any parent nodes $\mathbf{A}$ in Fig. 1 and $\mathbf{S}$ represent sensor nodes. The robot observes the human's behavior b_j and gathers the sensor information $\mathbf{d}_i$ at the same moment. Let $\mathbf{v}[t]\epsilon\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n\}$ be the observation of the sensor at time t. Let $o[t]\epsilon\{b_1, b_2, \ldots, b_m\}$ be the observation of the user's behavior at time t. The sensor observation vector is written as $\mathbf{V}[t] = \{\mathbf{v}[1], \mathbf{v}[2], \ldots, \mathbf{v}[t]\}$ and user instruction vector as $\mathbf{O}[t] = \{o[1], o[2], \ldots, o[t]\}$. Then, we can define data as $\mathbf{D}[t] = \{\mathbf{V}[t], \mathbf{O}[t]\}$. Let N be the number of observed data, N_j the number of observations of behavior b_j, and n_{ij} the number of observation of data when d_i is observed with b_j. One of the simplest calculations based on the observation is

$$P\,(\mathbf{d}_i|b_j) = P\,(S = \mathbf{d}_i|B = b_j) = \frac{n_{ij}}{N_j}, \qquad (2)$$

$$P\,(b_j) = \frac{N_j}{N}, \qquad (3)$$

A problem arises with this simple calculation when the data $\mathbf{D}[t]$ is continuously input during the observation. Suppose that the propositions of behavior b_1 and b_2 are set in the behavior node. When a rare but important operation b_2 is observed even though N_2 is smaller than N_1 the prior probability $P\,(B = b_2)$ is close to 0 while $P\,(B = b_1)$ is close to 1. The problem arises because the prior probability is calculated using the numbers of observations. We propose an approach in which the important observation is selected on basis of the change in the degree of confidence. When the change in confidence in two consecutive time steps is small, this situation is regarded as familiar; the experience data is considered insignificant to be discarded. In contrast, when the robot detect a large change in confidence in two consecutive time steps, this situation is considered unfamiliar; the experience data is considered significant to be accepted. The next section discusses an algorithm to distinguish the above two cases.

2.3 Proposed method

We use a Dirichlet distribution to evaluate the significance of data based on changes in the degree of confidence. A m-directional Dirichlet distribution for $x = \{x_1, x_2, \ldots, x_m\}$, is given by

$$f_d\,(\mathbf{x}; \alpha_1, \ldots, \alpha_m) = \frac{1}{Z}\prod_k x_k^{\alpha_{k-1}}, \qquad (4)$$

where,

$$Z = \frac{\prod_{k-1}^{m} \Gamma\left(\alpha_k\right)}{\Gamma\left(\sum_{k-1}^{m} \alpha_k\right)}, \tag{5}$$

is a normalization factor, Γ is the gamma function and the, parameters α_m are assumed to be positive. The Dirichlet distribution parameters are expressed in terms of observations of different behaviors for example $\alpha_1 = 1 + N_1$. The system increases one Dirichlet parameter α_1 by observing behavior b_1. When α_1 becomes larger than the other Dirichlet parameters, the peak of the distribution moves within a small area at the end of corresponding variable as shown in Fig. 2.

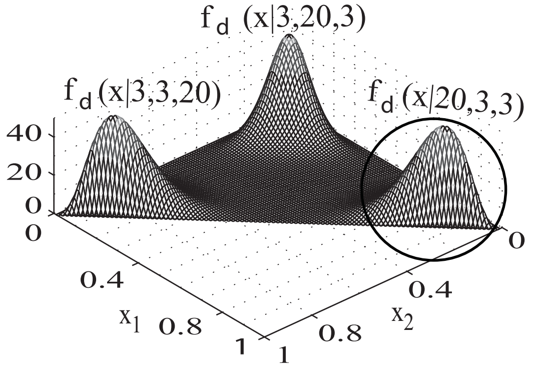

Figure 2: Dirichlet density functions with peak moved to the corresponding parameter

The system calculates the degree of confidence before and after the current observation. Confidence at time t is calculated as

$$C_t = \int_{\Delta} f_d\left(\mathbf{x}; \alpha^t\right) d\mathbf{x}, \tag{6}$$

where $f_d\left(\mathbf{x}; \alpha^t\right)$ is the Dirichlet distribution at time t and Δ represents area of integration where the peak of the distribution is moved like the area inside the circle in Fig. 2. The change in the two degrees of confidence can be regarded as the importance of the observation to the learning process. To evaluate the significance of the observation data, the criteria

$$E = C_t - C_{t-1}, \tag{7}$$

is calculated. Data $\{\mathbf{V}_i[t], \mathbf{O}_i[t]\}$ are accepted when $E \geq \theta$, and data are discarded when $E < \theta$. θ should be set according to required rapidness of the learning because significance of data can be controlled with θ. For example, an application with a very high input frequency will likely have a different threshold (a lower one) from one with a very low input frequency (relatively higher) for adapting to the user's new preference. For rapid adaptation, the area and threshold should be empirically determined by experimentation.

3 Experiment and Result

3.1 Experimental Setup

We developed a teaching and learning system in a virtual environment that incorporated our concept. The environment, as shown in Fig. 3 was prepared using webot real-time simulation software. The environment had an enclosed area of 8 [m]× 8 [m]. A static square obstacle whose size was 1 [m]× 1 [m] was placed inside the area. The user interface consisted of a lever joystick and the user controlled the robot by using it. We taught two policies to the robot, avoid and approach, in the field. In the experiment, we used a Bayesian network consists of eight distance sensor nodes and a behavior node as in Fig. 4.

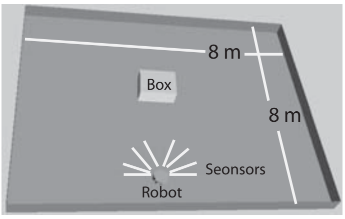

Figure 3: Virtual experimental environment

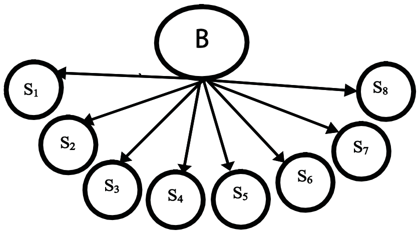

Figure 4: A Bayesian network used in the experiment

The robot model had eight front laser distance sensors (S_i,$i = 1, 2, \ldots, 8$) mounted on the front to measure the distance to obstacles along a horizontal line parallel to the floor. Joystick inputs were translated into discrete instructions by using a predetermined threshold. We found [8] that area of integration was inversely proportional and threshold value was directly proportional to the time required reach the discarding criteria respectively. Therefore we set the area of integration to the maximum non-overlapping area and the threshold to 1.0×10^{-6}.

The user can teleoperate the robot at any time and halt operation temporarily for changing robot orientation in the virtual environment. When user do not operate the robot, it operates automatically with it's own degree of confidence for behavior node. Previously we have shown that our algorithm can adapt to the user preference rapidly [8]. In that paper we have shown that robot adapted to user preference like go forward, turns left or turn right. In this experimental setup user policy correspond to robot behavior, avoids and approach. Avoid policy is accomplished by going forward when there is no obstacle and turning left when there is an obstacle. Approach policy is accomplished by going forward when there is no obstacle and approaching the obstacle when there is one.

3.2 Experimental Results

The user first taught avoid policy twice. The user then changed the policy and approached the obstacle. Fig. 5 shows the changes in probability of degree of confidence during teleoperation. When the user changed the policy from avoid to approach around step 250 the system could override the previous policy just after step 300 and the robot could rapidly adapted to the new policy.

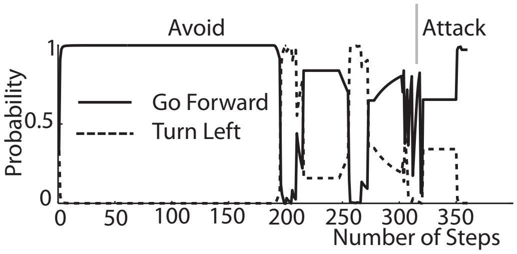

Figure 5: Probability of behavior during policy adaptation

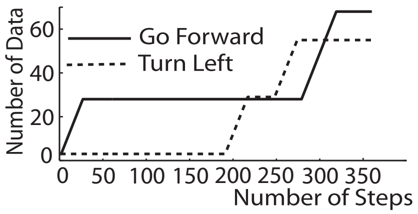

Figure 6: Number of data in the secondary database during policy adaptation

Fig. 6 shows the number of data evaluated as significant and kept in secondary database. The number of data in the secondary database is increased until the change in the degree of confidence was low for any user behavior. Flat part represents when data is evaluated as insignificant and discarded or the robot operate automatically. Here we observe that data is kept for one go forward and two turn left behavior for avoid policy. And when policy is changed the system accepted data for approach and overridden policy around after steps 300.

4 Conclusions

Experimental results show that proposed method can adapt to user's policy change based on significance evaluation using change in degree of confidence. The novel point of the method is that the policy adaption depends on the number of selected significant data rather than enormous amount of observed data. Currently, significance evaluation is done on the behavior node. We are considering significance evaluation for each sensor proposition of every sensor node. This will ensure that only significant sensor observation will be used for learning and that will make our system more robust.

References

[1] Widmer G. and Kubat M., *Learning in the Presence of Concept Drift and Hidden Contexts* in Machine Learning, **23**, 69-101, 1996.

[2] Klinkenberg R., *Learning Drifting Concepts: Example selection vs. Example weighting* in Intelligent Data Analysis, Special Issue on Incremental Learning Systems Capable of Dealing with Concept Drift, **8**(3), 281-300, 2004.

[3] Billsus D. and Pazzani M. J., *User Modeling for Adaptive News Access* in User Modeling and User-Adapted Interaction, **10**, 147-180, 2000.

[4] Chiu P. and Webb G., *Using Decision tree for Agent Modeling; Improving Prediction Performance* in User Modeling and User-Adapted Interaction, **8**, 131-152, 1998.

[5] T. Inamura, M. Inaba and H. Inoue, *PEXIS: Probabilistic Experience Representation Based Adaptive Interaction System for Personal Robots* in Systems and Computers in Japan, **35**(6), 98-109, 2004.

[6] T. Inamura, M. Inaba and H. Inoue, *User Adaptation of Human-Robot Interaction Model based on Bayesian Network and Introspection of Interaction Experience* in Proc. of Int'l Conf. on Intelligent Robots and Systems (IROS 2000), 2139-2144, 2000.

[7] M. Nicolescu and M. Mataric, *Learning and interacting in human-robot domains* in IEEE Transactions on System in Man and Cybernetics-Part A: Systems and Humans, **31**(5), 419-430, 2001.

[8] Tareeq S. M. and Inamura T., *A sample discarding strategy for rapid adaptation to new situation for Bayesian behavior learning* in Proceedings of the IEEE International Conference on Robotics and Biomemtics, 1950-1955, 2008.