Author: F. Ge, W. Moore, M. Antolovich and J. Gao

Title: Application of learning from demonstration to a mining tunnel inspection robot

Year: 2011

Conference Name: IEEE International Conference on Robot, Vision and Signal Processing (RSVP)

Conference Location: Taiwan

Publisher: IEEE

Date: 21-23 November, 2011

Abstract: Current research for Learning From Demonstration (LfD) seems to concentrate on the learning kernel. This paper outlines the need for a more useful variable selection technique using the training dataset. The paper presents a new training dataset selection method, called Information Extraction (IE). The application area is a complex task involving robot mining tunnel inspection, and IE is applied to the robot for this task. The Gaussian Mixture Model (GMM) is adopted to generate a learning curve utilized by a robot. The Gaussian Mixture Regression (GMR) is used to infer actions based on given states. After human demonstration, the robot can finish a pre-defined task independently.

DOI: http://dx.doi.org/10.1109/RVSP.2011.75

URL: http://researchoutput.csu.edu.au/R/-?func=dbin-jump-full&object_id=30430&local_base=GEN01-CSU01

Author Address: fge@csu.edu.au/ / mantolovich@csu.edu.au/ jbgao@csu.edu.au/

Application of Learning from Demonstration to a Mining Tunnel Inspection Robot

Fenglu Ge, Wayne Moore, Michael Antolovich, Junbin Gao
School of Computing and Mathematics
Charles Sturt University
Bathurst, Australia
Email: fge@csu.edu.au

***Abstract*—Current research for Learning From Demonstration (LfD) seems to concentrate on the learning kernel. This paper outlines the need for a more useful variable selection technique using the training dataset. The paper presents a new training dataset selection method, called Information Extraction (IE). The application area is a complex task involving robot mining tunnel inspection, and IE is applied to the robot for this task. The Gaussian Mixture Model (GMM) is adopted to generate a learning curve utilized by a robot. The Gaussian Mixture Regression (GMR) is used to infer actions based on given states. After human demonstration, the robot can finish a pre-defined task independently.**

***Keywords*-LfD; IE; GMM; GMR;**

I. INTRODUCTION

The inspection of muck piles for oversized rocks is of interest to the authors. It is possible for a worker to be assigned to this task but the nature and speed required for the inspection regime that will be used makes the use of a human inspector problematic. Previously, humans were hired to inspect mining tunnels, which required special training and specialised equipment [1]. To inspect muck piles, an inspector needs to periodically go to many (>400) underground positions, carrying a monitoring device in complex mining tunnels. Load Haul Dump (LHD) trucks travel around in a harsh environment, which will increase the probability of human-truck collisions and influence the inspection efficiency. To overcome these problems, robots have been introduced into mining tunnel inspection [2], which can reduce human error and increase efficiency. However, the robot is still teleoperated, and the teleoperator needs to co-locate with the robot for all possible inspections. Another potential solution is to programme a robot for a task in a slow time-varing environment. This has the drawback of adapting to unpredictable real environments, which leads a robot to stop or to be damaged because of lack of environment knowledge.

In an effort to overcome the above issues, 'Robot Learning' is applied to a robot for the purpose of mining tunnel inspection. Robot Learning is a technique allowing a robot to learn a task from demonstration, without pre-programming for each example of different tasks or environments. Therefore, Robot Learning is useful for a robot to perform a task, which releases a teleoperator from fully controlling the robot. Learning from Demonstration (LfD) and Reinforcement Learning (RL) are two major methods in the field of Robot Learning [3, 4]. RL has the advantage of making a robot adapt to an environment autonomously. However, it's difficult to give a suitable definition of reward [4], which is very important to guide the robot to achieve the most appropriate behaviour for the task to be learned. Moreover, building a policy [4] for RL also requires state information of the ambient environment so as to receive rewards, which costs time and is computationally intense in the real world. LfD [3] is a supervised technique, where a demonstrator can teach and guide a robot's behaviours during a training period . Compared to RL, LfD is a faster learning method for a robot since it doesn't need an unpredicted reward, saving time and computation.

The first stage of LfD is training dataset selection. Researchers copy environment data directly [5], or use filtered data [6] as a training dataset. In [6], the selection of time in training data makes a task dependent on critical time correction. In [5], environment data is used as a training dataset directly, but this is not suitable for a complex task described in this paper. This can lead to the same states corresponding to different actions. Therefore, training data selection and generalization is important for LfD. In this work, IE is introduced to give one solution for training dataset selection.

In this paper, LfD combined with IE is adopted for use in robot mining tunnel inspection. Section 2 will describe the system structure and theoretical analysis in this task. Section 3 will give experimental results and discussions. Section 4 will give a conclusion.

II. SYSTEM CONSTRUCTION AND THEORY

In this task, training data includes the following variables:

1) flag—an indicator of action change
2) lsp—linear speed of a robot.
3) asp—angle speed of a robot.
4) fd—distance to the nearest front object.
5) ld—distance to the nearest left object.
6) rd—distance to the nearest right object.

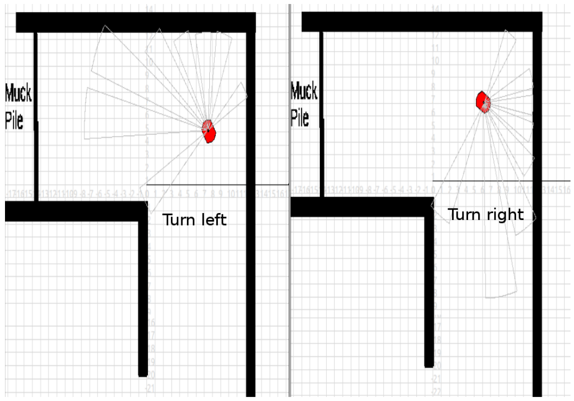

Figure 1: The robot turning direction varies given similar states.

A. Platform Setup

In this experiment, a P3-DX robot was used to acquire data at a sample rate of 1Hz. Two mini-desktops are on the platform. one supplies 3D video to a remote teleoperator. Another one receives commands from the remote teleoperator, and controls the robot.

B. Information Extraction

IE is a process which finds hidden information from an original dataset firstly, then combines the hidden information and original dataset together to generate a new training dataset. Previously, researchers copied the environment data directly [5], or use the filtered original dataset [6] as the training dataset.

In [5] LfD is used in a corridor navigation task, where three variables fd, ld, rd are chosen to deduce lsp and asp. However fd, ld and rd are not enough to infer lsp and asp in this mining task, which can generate different actions for similar states. In the left picture of Fig.1, the robot is turning left, while the robot is turning right in the right picture of Fig.1.

In [6, 7], time index is chosen as part of the training dataset. Therefore, every time index corresponds to a certain action. If the robot meets any exceptions, time index is still being counted when it is dealing with that exception, which will affect subsequent behaviours of the robot. Moreover, the robot can not easily extrapolate its knowledge learned because of the critical dependence of time.

To overcome the difficulties mentioned above, the training dataset needs to be developed so as to make a task portable to similar domains. Besides the original dataset collected by sensors, different information can be extracted from it to meet different task purposes.

In this task, a flag is generated according to changes in fd, ld and rd. With the introduction of a flag, a state leads to only one action. Fig.2 shows the relationship between absolute distance vector value and asp, where the distance vector is composed of fd, ld and rd. As shown in Fig.2, a state can lead to different turning directions. Fig.3 shows a state can lead to only one action with the introduction of IE. Three groups are displayed in Fig.3, where the top group is left turn, the middle is forward/stop and the bottom is right turn.

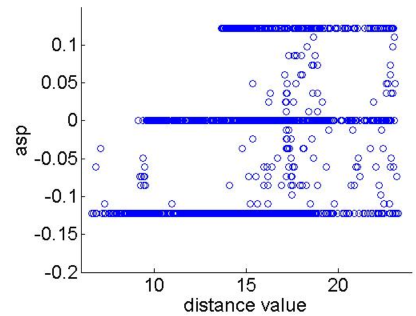

Figure 2: Distance vector VS asp without IE

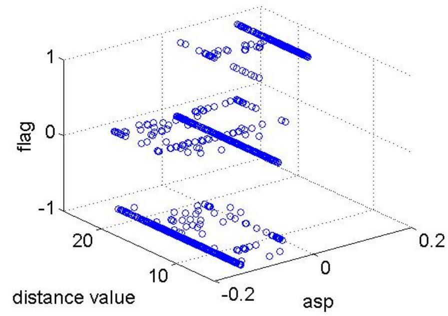

Figure 3: Distance vector VS asp with IE

C. Theoretical Base

To perfom LfD, the robot needs to summarize a classifier during the learning process, and the following steps are required:

1) K-means: K-means is used to separate input data into K clusters based on Euclidean Distance. The cluster centers for every cluster are also retrieved for future use.

2) EM Initialization: Expectation-Maximization (EM) [8] is initialized to get the inital parameters of a GMM. In this task, the input dataset is modelled by a mixture K components, where K is the number of clusters (actions).

$$p(X_i) = \sum_{k=1}^{K} p(k)p(X_i|k) \tag{1}$$

where $p(k)$ is the prior Probability Density Function (PDF), X_i is a d-dimensional vector and $p(X_i|k)$ is the posterior PDF.

In this paper,

$$X_i = [flag, fd, ld, rd, lsp, asp] \tag{2}$$

For every cluster, a gaussian model is used as PDF of the data distribution in a cluster, which can be denoted as follows:

$$\begin{aligned} p(X_i|k) &= N(\mu_k, \Sigma_k) \\ &= \frac{1}{\sqrt{(2\pi)^d |\Sigma_(k)|}} e^{-\frac{1}{2}(X_i-\mu_k)^T \Sigma_k^{-1}(X_i-\mu_k)} \end{aligned} \tag{3}$$

μ is a d-dimensional vector of X_{set}.Σ is $k \times k$ covariance matrix of X_{set}.

$$X_{set} = (X_i)_{i=1}^{M} \tag{4}$$

In this paper, X_{set} are the data collected by a robot in one task.

3) EM: Parameters of GMM are calculated by EM After the step of initialization, E-step and M-step [9] are used to retrieve μ and Σ.

E-step:

$$p_{k,i}^{t+1} = \frac{p_k^t N(X_i|\mu_k^t, \Sigma_k^t)}{\sum_{j=1}^{K} p_j^t N\left(X_i|\mu_j^t, \Sigma_j^t\right)} \tag{5}$$

M-step:

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{M} p_{k,i}^{t+1} X_i}{\sum_{i=1}^{M} p_{k,i}^{t+1}} \tag{6}$$

$$\Sigma_k^{t+1} = \frac{\sum_{i=1}^{M} p_{k,i}^{t+1} (X_i - \mu_k^{t+1})(X_i - \mu_k^{t+1})^T}{\sum_{i=1}^{M} p_{k,i}^{t+1}} \tag{7}$$

4) Gaussian Mixture Regression (GMR): GMR is an algorithm for data reconstruction [9]. It is used to estimate a variable Y given X on the basis of a series of previous observations $\{X, Y\}$. In this task, the previous observations are the data acquired by a robot, which are $\{flag, fd, ld, rd, lsp, asp\}$. The target is to estimate the new output $\{lsp, asp\}$ given new input $\{flag, fd, ld, rd\}$ in the basis of previous observations. Here we define:

$$X_{in} = \{flag, fd, ld, rd\} \tag{8}$$

$$X_{out} = \{lsp, asp\} \tag{9}$$

For every Gaussian component k, we define:

$$\mu_k = \{\mu_{in,k}, \mu_{out,k}\} \tag{10}$$

$$\Sigma_k = \begin{pmatrix} \Sigma_{inin,k} & \Sigma_{inout,k} \\ \Sigma_{outin,k} & \Sigma_{outout,k} \end{pmatrix} \tag{11}$$

and the following equation can be inferred:

$$p(X_{out,k}|X_{in,k}) = N(\mu_{GMR}, \Sigma_{GMR}) \tag{12}$$

$$\mu_{GMR} = \mu_{out,k} + \Sigma_{outin,k} (\Sigma_{inin})^{-1} (X_{in} - \mu_{in,k}) \tag{13}$$

$$\Sigma_{GMR} = \Sigma_{outout,k} - \Sigma_{outin,k} (\Sigma_{inin,k})^{-1} \Sigma_{inout,k} \tag{14}$$

Therefore, it can be deduced that

$$p(X_{out}|X_{in}) = \sum_{k=1}^{K} \gamma_k N(\mu_{GMR}, \Sigma_{GMR}) \tag{15}$$

$$\gamma_k = \frac{p(k)\, p(X_{in}|k)}{\sum_{i=1}^{K} p(i)\, p(X_{in}|i)} \tag{16}$$

Thus, the value of X_{out} can be estimated by X_{in} via following equations:

$$X_{out} = \sum_{k=1}^{K} \gamma \mu_{GMR} \tag{17}$$

$$\Sigma_{outout} = \sum_{k=1}^{K} \gamma^2 \Sigma_{GMR} \tag{18}$$

III. Experiment and Results

The experiment with a P3-DX robot can be summarized as follows:

1) a teleoperator controlled a robot to run through a training environment in Fig.4, teaching a robot to finish a task. The robot used its built-in sonar array to record data at the same time.
2) IE was applied to the raw data to form a new training dataset. The robot used the new training dataset via algorithms to generate a classifier.
3) The robot tried to finish the task independently. At every time point, it recorded flag, fd, ld, rd and put them into the classifier. The classifier processed the input to find an output with the highest probability, where the output was an action the robot needed to take.
4) If the robot didn't work properly, the task would be interrupted by a supervisor. Extra training was performed at the spot where the robot made mistakes, and new training data was added to the previous training data. Step 2 and Step 3 repeated until the robot could finish the task independently.

During the training period, a teleoperator controlled a robot to run through a tunnel shown in Fig.5. Here, Action 1, 3, 6, 8 are forward. Action 4 is stop. Action 2 is turning left at 90 degrees. Action 7 is turning right at 90 degrees. Action 5 and 9 are U turn (turn right at 180 degrees). The robot stopped before a muck pile for a while, then went back to the starting point. Sonar was used to collect experimental data. Due to the noise and the interference of the environment in practice, the robot took average of 10 consecutive readings

Figure 4: Robot training environment

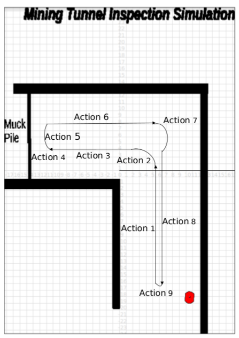

Figure 5: Running route of the robot during training period.

as input data. In this experiment, the robot had 5 actions, forward, turn left, stop, turn right and U turn. The forward action made the robot move at a linear speed of 1.5 m/s, while the angle speed was 0 deg/s. The turning actions turned the robot with an angle speed of 7 deg/s, while linear speed was 0 m/s. The robot took more than 9 actions to complete one loop of that domain.

In this experiment, off-line learning [3] was used to train the robot. 610 samples were used to initialize GMM, where a sample is a vector defined in (2). 1271 samples were used to update GMM. Fig.6 shows the route when the robot performed the task independently.

IV. Conclusion

In this paper, LfD combined with IE is used to encode environment data collected by a robot. The robot can perform

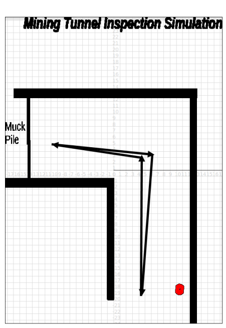

Figure 6: True route when the robot performed a task independently

a task via a GMM learning kernel and update the kernel with the help of teleoperation. The learning kernel can be applied to similar domains like the one in Fig.6. Our results show the learning kernel is robust. In the future, the learning kernel will be improved so as to increase the accuracy [10,11] and the speed of learning. Multiple robots will also be introduced and taught simultaneously in order to increase the teaching efficiency.

Acknowledgment

The authors gratefully acknowledge Newcrest Mining Ltd for funding this work.

References

[1] N. Gunningham, *Mine Safety: Law, Regulation, Policy*, Federation Press, Leichardt, 2007.

[2] J. C. González, S. Martínez, A. Jardón and C. Balaguer, *Robot-Aided Tunnel Inspection and Maintenance System*, 26th International Symposium on Automation and Robotics in Construction (ISARC), Austin, TX, USA, 2009.

[3] B. D. Argall, S. Chernova, M. Veloso and B. Browning, *A survey of robot learning from demonstration*, Robotics and Autonomous Systems 57(5): 469-483, 2009.

[4] L. P. Kaelbling, M. L. Littman, and A. W. Moore, *Reinforcement learning: A survey*, Journal of Artificial Intelligence Research, 4: 237-285, 1996.

[5] S. Chernova, M. Veloso, *Confidence-based learning from demonstration using Gaussian Mixture Models*, Proceedings of the International Conference on Autonomous Agents and Multiagent Systems, AAMAS07, 2007.

[6] S. Calinon, F. Guenter, and A. Billard, *On learning, representing and generalizing a task in a humanoid robot*, IEEE Transactions on Systems, Man and Cybernetics, Part B, Special issue on robot learning by observation, demonstration and imitation, 37(2): 286-298, 2007.

[7] S. Calinon, F. Dhalluin, E. Sauser, D. Caldwell and A. Billard, *Learning and reproduction of gestures by imitation: An approach based on Hidden Markov Model and Gaussian Mixture Regression*, IEEE Robotics and Automation Magazine, 17(2): 44-54, 2010.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum Likelihood from Incomplete Data via the EM algorithm*, Journal of the Royal Statistical Society, Series B, 39(1): 1-38, 1977.

[9] S. Calinon and A. Billard, *What is the teachers role in robot programming by demonstration? - Toward benchmarks for improved learning*, Interaction Studies, 8(3): 441-464, 2007.

[10] J. Li, S. Chu, J. H. Chang, and J. Pan, *Discriminant feature fusion strategy for supervised learning*, in International Conference on Intelligent Information Hiding and Multimedia Signal, California, 301-304, Dec, 2006.

[11] D. Y. Huang, C. J. Lin, and W. C. Hu,*Learning-based Face Detection by Adaptive Switching of Skin Color Models and AdaBoost under Varying Illumination*, Journal of Information Hiding and Multimedia Signal Processing, 2(3): 204-216, July, 2011.