Probabilistic Robot Localization and Situated Feature Focusing

Yoichi Motomura*, **Nikos Vlassis****, **Ben Krose****

*Electrotechnical Laboratory,
Tsukuba, Ibaraki 305-8568, JAPAN
motomura@etl.go.jp
**RWCP Autonomous Learning function SNN, University of Amsterdam,
Kruislaan403, Amsterdam, 1098 SJ, The Netherlands
vlassis@wins.uva.nl, krose@wins.uva.nl

Abstract

Robot localization, i.e., the task of recognizing the current position of the robot from sensor inputs is an essential problem for autonomous mobile robots. In this paper, we discuss the localization problem through probabilistic models, information theoretic criteria, and statistical learning. When we use some variety of sensors or high dimensional inputs like image pixels, decreasing first their dimensionality, or extract features, is necessary for making the data tractable. We will show popular feature extraction methods for localization and some properties of them. After feature extraction we can construct position estimation probabilistic models by regression. By probabilistic modeling, the information theoretic meaning of a feature extraction method becomes clearer. We introduce a mutual information-based criterion to evaluate the feature set, and compare this criterion with Kullback Leibler divergence and the average Bayesian localization error. In general, the evaluation result of the feature extraction depends strongly on the particular region of the environment. A feature performing well in a local region may not be good for the other local region. For an entire environment, an appropriate feature should be selected according to the corresponding situation. We call this idea *situated feature focusing* that select feature extraction modules and local regression models. This approach can be realized by Bayesian networks to estimate possibility of current situation and the mixture of experts which is the combination of various feature extraction.

1 Introduction

For intelligent autonomous robots which move in the real world, a localization task to identify the robot's position is essentially important. Matching the sensory inputs with a deterministic environment model has the drawback that it is not so robust for noisy and uncertain sensors and also because it is difficult to describe a complete model a priori. To overcome such problems, probabilistic modeling is studied as a promising approach [1, 2, 3, 8, 9]. In this framework, training the model to estimate locations from sensor inputs can be regarded as probabilistic regression.

Many sensors or camera images are necessary to get enough information for practical localization, thus decreasing the dimensionality internally is necessary for making a model tractable. This preprocessing corresponds to the feature extraction scheme of pattern recognition. Good feature extraction can decrease the redundancy of sensor inputs and allow invariance for irrelevant fluctuations of the signals.

Therefore, we use the following two steps for the robot localization problem: (1) project into relatively low-dimensional internal vector from raw sensor signal vector, and (2) apply probabilistic regression modeling from the robot's position to the internal feature vector. In earlier works, as a projection in (1) the principal component analysis (PCA) [4, 5, 9, 8], neural networks [2] and human designed features (specific mark, color, etc.) have been studied in this context. However there are practical problems, e.g., that neural network learning requires long time to converge, PCA as a linear projection does not have enough flexibility to approximate complex target environments, and the human designing features are effective for specific only environments and not for general ones. The important issue is not only obtaining better projection but also specifying criteria to select better projections, since it is not obvious what kind of projection is desirable.

In this paper we study a probabilistic modeling method for robot localization and investigate the use of a mutual information-based criterion for extracting better features. By probabilistic modeling, the information theoretic meaning of the feature extraction method becomes clearer. We also discuss differences among the mutual information based criterion, the Kullback Leibler divergence, and the average Bayesian localization error [2].

Finally, we maintain that good feature extraction should depend on the particular local region of the environment. For an entire environment, the robot should select appropriate feature sets for each current local region of the environment. For better feature extraction, we introduce the idea of 'situated feature focusing' involving local feature extraction modules and local regression models. This approach can be realized by Bayesian networks to estimate possibility of current situation and the mixture of experts which is the combination of various feature extraction. We believe that this is more consistent with the human active perceptual behavior to recognize the real environment.

2 Probabilistic modeling for localization

The robot observes a sensor vector z at a position x. The robot localization task is to correctly estimate x from z. For this task, an environment model $x = M(z)$ is required. The real environment is filled with uncertainty, then it is modeled as a conditional probability distribution $P(x|z)$, where x and z are random variables. This probability distribution is a non-deterministic realization of the model $x = M(z)$, and the probability of x can be regarded as the degree of belief that the robot is located in x.

In practical situations, two sources of the problem are encountered:

- Similar z may be obtained from different x (perceptual alias)
- z may change abruptly even if x changes slightly (non linearity)

From the former, a one to one mapping from z to x as a model is not adequate. From the latter, a non linear regression model may be necessary.

The sensor high dimensional vector z often includes noise and redundancy, so we use feature extraction as a preprocessing step to get a lower dimensional vector. This feature extraction is denoted by $y = f(z)$ (y is lower dimensional vector than z). If this feature extraction $y = f(z)$ is enough to distinguish x, we can estimate the robot locations from the probabilistic relationship between y and x.

The posterior probability of the robot position x given sensor reading z is described through a parametric model $P(y|x;\theta)$ and the Bayes theorem as

$$P(x|z) \approx P(x|y = f(z)) = \frac{P(y = f(z)|x;\theta)P(x)}{\int_x P(y = f(z)|x;\theta)P(x)dx}, \tag{1}$$

where θ is a parameter vector, and the prior distribution $P(x)$ is given. The integral operation in the denominator is marginalization for all possible x.

$P(x|y)$ may be multi-modal, however, by modeling as eq.(1), the parametric model $P(y|x;\theta)$ can be uni-modal, e.g., a single Gaussian distribution.

Thus, the environment modeling proceeds as follows:

- First choose appropriate feature extraction projection $y = f(z)$.
- Then estimate θ for the parametric model $P(y|x;\theta)$.

After the modeling steps, we can get an estimate of the robot's position x which maximizes the quantity

$$P(x|y = f(z)) \propto P(y = f(z)|x;\theta)P(x), \tag{2}$$

when the sensor reading is z.

3 Regression models

The modeling scheme described in the previous section can be established by statistical learning from a data set $D \equiv \{x_i, z_i\}, (i = 1, \cdots)$ obtained in the environment by the robot. This can be regarded as regression. In this section, we show two regression models combined with feature extraction for probabilistic robot localization.

3.1 PCA regression

Principal Component Analysis (PCA) is known as the method that gives lower information loss and lower dimensional vectors which are uncorrelated and orthogonal with each other. The projections obtained by PCA are linear combinations which maximize the variance of $\{f_{PCA}(z_i)\}, (i = 1, \cdots)$.

After applying PCA, we can use a parametric model with a single Gaussian,

$$P(y|x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} exp(-(g(x;\theta) - y)^2/2\sigma^2). \tag{3}$$

where, $g(x;\theta)$ and σ^2 denote a non linear function with parameter θ and the variance, respectively. The parameter θ in eq.(1) is decided by regression from the data set $D_f \equiv \{x_i, y_i = f_{PCA}(z_i)\}$. In [8], a generalized linear model was used for $g(x;\theta)$, giving rise to PCA regression. In this model, $y = f_{PCA}(z)$ is a linear deterministic function.

3.2 Neural networks

Neural networks have been applied to feature extraction under a Bayesian probabilistic formulation, an approach called Bayesian landmark learning [2]. The feature extraction scheme is modeled by a conditional probability $P(y|z)$ as the output of a feed forward neural network $f_{NN}(z)$. The estimation of the posterior probability of position x given z is given by the Bayes' theorem and the Markov assumption[1] .

The neural network $f_{NN}(z)$ is trained by a gradient descent argolithm so as to minimize the average Bayesian localization error (discussed in the next section), giving rise to 'useful' features for robot localization.

4 Criterion for localization

For the principal component regression model, f_{PCA} is decided by PCA (maximizing the variance) and θ is decided by regression (least mean squares fitting). However, what seems to lack is an information theoretic meaning of PCA that would, for example, shed light to which principal components should be selected as good features for localization. Therefore, we would

[1] Sensor readings are conditionally independent of previous readings and the action of the robot if the robot's location is unknown.

like to understand the general desirable property behind feature extraction and the estimation of probabilistic models in this context. For this purpose we introduce a mutual information-based criterion and discuss about the properties of the PCA regression model under this criterion.

4.1 Kullback Leibler divergence

Kullback Leibler (KL) divergence is used for measuring the distance between probability distributions. If we imagine the data set is generated from an unknown true distribution, then the KL divergence between the true distribution and an estimated distribution modeled by eq.(3) is introduced naturally. We denote the true distribution as $P^*(x,y)$ in the $x-y$ space. The estimated distribution of eq.(3) is $P(x,y) = P(y|x;\theta)P(x)$. Therefore, the KL divergence is described as

$$\begin{aligned} & \int_x \int_y P^*(x,y) \log \frac{P^*(x,y)}{P(y|x;\theta)P(x)} dydx \\ = & \int_x \int_y P^*(x,y) \log P^*(x,y) dydx \\ - & \int_x \int_y P^*(x,y) \log P(y|x;\theta)P(x) dydx. \quad (4) \end{aligned}$$

The first term corresponds to the negative entropy defined by D and $y = f(z)$. The second term corresponds to the log-likelihood of the training set which we denote by $L_g(\theta)$.

4.2 Mutual information-based criterion

The first term in the definition of the KL divergence above says that this criterion prefers larger variance when the second term (log-likelihood) remains constant. If we apply the KL divergence criterion for robot localization, this implies choosing projections which give larger variance for equal likelihood. This property may make the estimation result ambiguous. For preciser estimation, the selected projection should give lower entropy, and this implies an alternative criterion for the robot localization: the ambiguity of the estimation is evaluated by the expected entropy of $P(x|y = f(z))$ which is denoted by $E_y[H(x|y)]$ and can be calculated as

$$\int_y \int_x -P^*(x|y) log P^*(x|y) dx P^*(y) dy, \quad (5)$$

where

$$\begin{aligned} P^*(y) & \equiv \int_x P^*(x,y)dx \quad (6) \\ P^*(x|y) & \equiv P^*(x,y)/P^*(y). \quad (7) \end{aligned}$$

Then $E_y[H(x|y)]$ can be decomposed by Bayes' theorem as

$$E_y[H(x|y)] = E_y[H(y|x) + H(x) - H(y)]. \quad (8)$$

Since $H(x)$ is completely determined by D, selecting $y = f(z)$ affects only the terms $H(y|x) - H(y)$. Using this instead of the first term of eq.(4), we get a new criterion for robot localization as

$$E_y[H(y|x) - H(y)] - L_g(\theta) \quad (9)$$

The model minimizing this criterion should give a correct estimation of the position of the robot. Minimizing eq.(8) means selecting y which maximizes the mutual information $H(x) - H(x|y)$, thus this criterion is also consistent with our intuition.

When we calculate the entropy, histograms of D are necessary. If the model can fit to D well, the entropy can be obtained by analytical calculation about the model. In the case of PCA regression eq.(3), $H(x|y)$ becomes constant with respect to σ. The m-th selected projection $y^{(m)}$ by PCA is ordered so as to maximize the variance, i.e., if $m < n$ then $H(y^{(m)}) > H(y^{(n)})$. This result says that feature extraction by PCA can minimize the first term in eq.(9).

Next we must take into account the second term in eq.(9) which requires the unknown true $P^*(x,y)$. Substituting this, we calculate the log-likelihood from the data set D, deriving the second term of eq.(9) with eq.(3)

$$\begin{aligned} L_g(\theta) & \propto \sum_i \log P(y = f(z_i)|x_i;\theta) \\ & = \sum_i \log \frac{exp(-(g(x_i;\theta) - f(z_i))^2/2\sigma^2)}{\sigma\sqrt{2\pi}} \\ & = \alpha - \beta \sum_i (g(x_i;\theta) - f(z_i))^2, \quad (10) \end{aligned}$$

where, α and β are constants decided by σ.

Accordingly, θ given by the least squares error fitting can optimize eq.(10). This means that the second term of eq.(9) is also optimized by the PCA regression model.

Finally, it has been experimentally shown [8] that the PCA regression model gives optimum estimation under the criterion of eq.(9) when regression after PCA is achieved perfectly. However, PCA just maximizes $H(y)$ but does not commit to $H(y|x)$. It rather relies upon the selected regression scheme. If the generalized linear model after PCA is not enough for perfect regression, i.e., $H(y|x)$ is not consistent with σ, then we need a more powerful feature extraction method for maximizing the mutual information [2].

4.3 Average Bayesian localization error

In [2] it was proposed an average Bayesian localization error as a criterion for learning landmarks with neural networks. When x^* denotes the true position and x denotes the estimation position, the estimation error is

[2] Instead of PCA, our group is also investigating the Projection Pursuit method [11].

measured by $||x - x^*||$. For a single x^*, the estimation error of $P(x|y)$ is defined as

$$Err(x^*) = \int_x \int_y ||x - x^*|| P(y|x^*) P(x|y) dx, \quad (11)$$

where $P(x|y)$ denotes the posterior probability distribution given y and $P(y|x^*)$ denotes the likelihood of observing y from x^*. This formula defines an average position estimation error for all possible y from x^*. Since the model is defined by $P(y|x;\theta)$, the above is transformed by Bayes' theorem as

$$Err(x^*) = \int_x \int_y ||x - x^*|| P(y|x^*) P(y|x) P(x) P^{-1}(y) dy dx, \quad (12)$$

$$P(y) = \int_x P(y|x) P(x) dx. \quad (13)$$

Moreover, we can imagine the average error for all possible true position x^* which gives

$$\begin{aligned} Err = & \; avg_{x^*} Err(x^*) \\ = & \int_{x^*} Err(x^*) P(x^*) dx^* \\ = & \int_{x^*} \int_x \int_y ||x^* - x|| P(y|x^*) \\ \cdot & \; P(y|x) P(x) P(x^*) P^{-1}(y) dy dx dx^* \quad (14) \end{aligned}$$

This is the average Bayesian localization error[2].

In [2], y is restricted to binary variables and it is recognized as a kind of a landmark. Then, training a neural network to optimize the above error realizes an automatic landmark detection or feature extraction mechanism.

The computation of the average Bayesian localization error requires marginalization over three variables x^*, x, and y, with complexity $\mathcal{O}(|y||x|^2)$, ($|x|$ denotes the size of the x space). Furthermore, since the method requires the calculation of the gradient for learning iteration, the total computational cost may be a problem [3].

5 Situated feature focusing

In the previous sections, some probabilistic methods for obtaining good feature extraction are discussed. The PCA regression model achieves features by a globally linear projection $y = f_{PCA}(z)$, and the neural network approach achieves a generative non-linear projection $y = f_{NN}(z)$. Both of them are global, i.e., over the entire robot space, projections.

In a small region of the environment such a global feature extraction method can work well. However, in general, it is not so easy to scale it up to larger and more complex environments, because even if the projection is non-linear for the input space, the projection is the same in different locations. In practical cases it is easy to find out the most relevant sensor in different regions of the robot space. For a real environment, an appropriate feature should be selected according to the corresponding situation.

[3] In [2] it is reported that learning required 12 and half hours on a Pentium Pro 200MHz

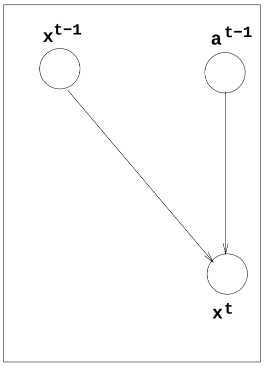

Figure 1: Bayesian network expecting region

One approach is combining local projections which depend on particular locations. Each local model assumes responsibility for relevant features in its own situation which could be, e.g., a limited region of the whole environment, a corridor, etc. Thus, we can obtain each feature extraction projection $f_c(z)$ for each region c. For example, PCA and neural network learning can be applied to yield f_c from the data set D_c sampled from a respective bounded region c.

In order to select an appropriate projection f_c, we have to estimate the current situation c. This can be modeled as a conditional probability $P(x^t \in c|x^{t-1}, a^{t-1})$, where x^{t-1} is the previous position of the robot, x^t is the next position, c is a region and a^{t-1} is an action at position x^{t-1}. This conditional probability has been implemented by a Bayesian network (Figure 1) and trained from examples taken by the real mobile robot[7]. We call this idea *situated feature focusing*, and the model for it can be constructed as the following.

$$P(x^t|z, x^{t-1}, a^{t-1}) = \sum_c P(y = f_c(z)|x^t) P(x^t \in c|x^{t-1}, a^{t-1})/Z (15)$$

where Z is a normalizer. The above can be regarded as a mixture of experts model [6]. By this method we can focus on the extraction of relevant combinations of features for a particular situation. This property bears resemblance to the human cognitive behavior (so called focus of attention).

6 Conclusions

We derived a probabilistic formulation of the robot localization problem and discussed the use of a mutual information-based criterion for feature extraction. The justification of the PCA regression model and also its limitations becomes clearer by this criterion. Neural network learning is also possible by this criterion instead of optimizing the average Bayesian localization error. In this case, it is expected that the computational cost will be decreased as mentioned in the section 4.3. For much better feature extraction in the entire environment we introduced the idea of situated feature focusing. Learning the mixture of local models with Bayesian networks and their experimental evaluation will be the next important issue.

Acknowledgements

This work is supported by the Real World Computing Project, and RWCP provided us with the autonomous robot database MEMORABLE. The authors would like to thank Roland Bunschoten, Bert Kappen, David Barber, Wim Wiegerinck, Toshihiro Matsui, Hideki Asoh, and Isao Hara for valuable discussions.

REFERENCES

[1] T.Dean and M.Wellman, "Planing and control", Morgan Kaufmann,CA ,1991.

[2] S.Thrun, "Bayesian landmark learning for mobile robot localization", *Machine Learning*, 33(1),1998,.

[3] S.Oore, G.E.Hinton and G.Dudek, "A mobile robot that learns its place", *Neural Computation*, 9, 1997, pp.683-699.

[4] S.Maeda, Y.Kuno and Y.Shirai, "Active navigation vision based on eigenspace analysis", *proc. of Int.conf. on Intelligent Robots and Systems*, 1997.

[5] J.L.Crowley, F.Wallner, B.Schiele, "Position estimation using principal components of range data", *the Proc. of int. conf. on Robotics and Automation*,1998.

[6] M.Jordan, "Hierarchies of adaptive experts", *Advances in Neural Information Processing*, 4, San Meteo CA Morgan Kaufmann, 1992.

[7] Y.Motomura, I.Hara, H.Asoh, and T.Matsui, "Bayesian Network that Learns Conditional Probabilities by Neural Networks", *Progress in Connectionist-Based Information Systems, vol.1*, 1997, pp.584-587.

[8] N.Vlassis and B.Krose, "Robot environment modeling via principal component regression", *proc. of Int.conf. on Intelligent Robots and Systems*, 1999,*to appear.*

[9] B.Krose and R.Bunschoten, "Probabilistic localization by appearance models and active vision", *proc. of Int.conf. on Robotics and Automation* 1999, *to appear.*

[10] N.Vlassis, Y.Motomura, and B.Krose, "An Information-theoretic localization criterion for robot map building", *ACAI '99*, 1999, *to apear.*

[11] N.Vlassis, Y.Motomura, and B.Krose, "Supervised Projection Pursuit with a Mutual Information Index", *submitted to NIPS'99*, 1999.