1999 年情報論的学習理論ワークショップ
1999 Workshop on Information-Based Induction Sciences (IBIS'99)
Izu, Japan, August 26-27, 1999.

Robot Localization における特徴選択のための相互情報量に基づく評価基準
Mutual Information Based Criterion for Feature Selection of Robot Localization

本村 陽一* Nikos Vlassis † Ben Krose ‡
Yoichi Motomura Nikos Vlassis Ben Krose

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. We focus on the localization problem as probabilistic modeling , information theoretic criteria, and statistical learning. When we use some variety of sensors or high dimensional inputs like camera image pixels, decreasing first their dimensionality, or extract features, is necessary for making the data tractable. After feature extraction we can construct position estimation probabilistic models by regression. How to select good feature is controversial issue. Many previous research used *principal component analysis.* In order to evaluate good feature selection, a mutual information-based criterion is proposed. Under this criterion, PCA regression model is justified when regression is achieved with small variance. However, in general, the robot's moving area being wider, a good feature for particular local region may not be good for the other local region by means of the criterion. For an entire environment, an appropriate feature should be selected according to the corresponding situation. As a solution for this problem, conditional mixture of PCA is proposed.

Keywords: Robot localization, Probabilistic model, Principal component analysis, Mixture model

1 はじめに

実環境で知的なタスクを行うための自律移動ロボットにとって, 不確実な要因に満ちた実環境の中で自己位置を同定することは本質的に重要なタスクである. ロボットが内部に環境のモデルを明示的に用意してこれとセンサからの情報と照合する方法ではしばしば起こるセンサのノイズや観測情報の変化や欠損に対して頑健でないことや, 実際の環境を完全に記述することが容易でないなどの問題がある. 実際の環境のこうした不確実性を取り扱うために確率モデルを用いる非決定的なアプローチが有効である [3, 4, 5, 9, 10]. また, ロボットが環境を認識するのに十分な情報を得るためには多くのセンサや, カメライメージなどが必要とされ, 入力情報の次元は比較的大きくなるため, 内部的な取り扱いを容易にするために次元を下げることが行われる. これはパターン認識における特徴抽出に相当しロボットがセンサ情報から特徴 (ランドマーク) を抽出し, より次元の低いコンパクトな表現にすることで, 信号の冗長性の低減, 不要な信号変化に対する不変性などが期待できる. そこでロボットの位置推定のためのモデルとして, まずセンサ信号からより低次元のベクトルを取り出す特徴選択と, これに対して場所を推定するための回帰の二段階の過程を考えることにする.

これまでの研究では, 特徴選択として線形の主成分分析 (PCA) を用いたもの [6, 7, 9, 10], 非線形写像であるニューラルネットを用いたもの [4, 5] や, ヒューリスティクスにもとづいて人間がデザインした特徴 (特定のマークや色など) を用いる例などがある. ただし, ニューラルネットの学習を用いたものでは学習に時間がかかったり, PCA は線形写像であるためモデルの自由度が少なく推定精度が不十分なことがあったり, ヒューリスティクスにもとづくものでは人的コストや, 環境固有の知識に強く依存するために汎用的でないなどの実際的な問題点がある. また特徴選択の良さだけを直接評価することは容易ではなく, 本タスクにおいてどのような評価基準が計算論的に妥当であるかが明らかではない.

*電子技術総合研究所, 〒 305-8568 つくば市梅園 1-1-4, 0298-54-5836, motomura@etl.go.jp, Electrotechnical Laboratory, 1-1-4 Umezono, Tsukuba, Ibaraki, 3058568, Japan
†RWCP Autonomous Learning function SNN, University of Amsterdam, Amsterdam, the Netherlands
‡RWCP Autonomous Learning function SNN, University of Amsterdam, Amsterdam, the Netherlands

本稿では Robot localization を確率的なモデルにより定式化し, まず PCA による特徴選択とガウスモデルによる回帰を使った例を見る. 次に特徴選択のための評価基準を検討し, 相互情報量に基づく基準について述べ, PCA 回帰モデルがこの基準のもとで正当化される条件を考察する. 最後に, 実際の応用にあたってロボットが移動する領域を拡大する場合に生じる問題点と, これを解決するための階層的なモデル, Conditional mixture of PCA についても述べる.

2 確率的モデル化

ロボットは環境中のある場所 x(例えば 2 次元ベクトル) において, 複数のセンサから入力されるベクトル z を得るものとする.

さて, ロボットが場所を認識する問題はあるセンサ入力 z を得た時に正しい x を推定することである. このためにロボットは学習によって環境に関するモデル $x = M(z)$ を獲得することが必要である. しかし前章で述べたような環境の不確実性を扱うためにこれを条件付き確率分布 $P(x|z)$ によってモデル化する. この場合の x, z は確率変数 (ベクトル) である. この確率分布は $x = M(z)$ についての非決定的出力でもあり, 確率値はセンサ入力を得た時のロボットが x にいる確信度として解釈することもできる.

ここで, 実際問題として (i) 遠く離れた場所からでも類似したセンサ値が得られる可能性, (ii) 比較的近い場所であってもセンサ値が急峻に変化する可能性の二点を考慮しなければならない. (i) から環境モデル M を z から x への一対一写像として表すことは適切でなく, また (ii) から非線形なモデルが必要な場合がある. さらにセンサ値 z はノイズや冗長性を持つことが多いので前処理として特徴選択を行い, より次元の低い表現に変換する. この特徴選択をある写像変換 f を用いて $y = f(z)$ と表す. さらに y と x の間を関係づける確率モデルと合わせて構成する (図 1).

センサ入力 z のもとでロボットの自己位置に関する事後確率は θ をパラメータとする条件付き確率分布 $P(y|x;\theta)$ とベイズの定理によって以下のように書ける.

$$P(x|z) = \frac{P(y = f(z)|x;\theta)P(x)}{\int_x P(y = f(z)|x;\theta)P(x)dx}. \qquad (1)$$

ここで $P(x)$ はロボットが位置する場所に関する事前確率分布で, 分母の積分は全ての取り得る x に関する周辺化である. 先に述べた (i) の理由から $P(x|y)$ は多峰性の分布である可能性があるが, 式 (1) のようにモデル化することで, パラメトリックモデル $P(y|x;\theta)$ として単峰性の分布 (ガウス関数) を使うこともできる.

図 1: Robot localization のためのモデル

環境のモデリングは

- 適切な特徴選択 $y = f(z)$ の選択,
- パラメトリックモデル $P(y|x;\theta)$ の θ の推定

が必要で, これをロボットが経験的に獲得したデータセット $D \equiv \{x_i, z_i\}, (i = 1, \cdots)$ からの統計的学習で行う.

f, θ が決定すると, センサ情報 z を観測した時,

$$P(x|z) \propto P(y = f(z)|x;\theta)P(x). \qquad (2)$$

を最大にする x をロボットの現在位置の推定値として得る.

3 PCA 回帰モデル

多次元のデータベクトルが多数与えられている時, できるだけ情報損失が少なく, より低次元の互いに無相関なベクトルに変換する手法が主成分分析 (Principal Component Analysis) として知られている. これは全てのベクトル z_i の分散共分散行列を Z とし, $Zw_j = \lambda_j w_j$ の固有値計算から得られた固有ベクトル w_j のうち固有値の大きい順に主要な成分を q 個だけ取り出して並べた行列 W によって,

$$y = f(z) = W^T Z = W^T(z - \bar{z}) \qquad (3)$$

という線形変換を構成するものである. これによる写像は互いに直交, 無相関で変換後の分散を最大にするような線形結合となる.

次に主成分分析を行った後の y についてガウス関数を使ったパラメトリックモデル,

$$P(y|x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} exp(-(g(x;\theta) - y)^2/2\sigma^2). \qquad (4)$$

を考える．ここで $g(x;\theta), \sigma^2$ はそれぞれ θ をパラメータとする非線形関数, ガウス関数の分散パラメータである．これらを定めた後で式 (1) におけるパラメトリックモデルの推定は θ を決定することである．g として例えば中心を固定した複数のガウスカーネルの線形結合とすると，その係数 θ は D に PCA を施した後のデータ $D_f \equiv \{x_i, y_i = f(z_i)\}$ による回帰 (最小二乗推定) でワンショットで求めることができる [9]．これを PCA 回帰モデルと呼ぶ.

4 特徴選択のための評価基準

PCA 回帰モデルにおいて, f の決定は PCA の分散最大基準により行われ, θ は D_f についての回帰 (最小二乗基準) によって決定される．しかし, このように定められた基準の理論的妥当性は明らかではない．またロボットの場所同定問題のための特徴 f の選択に関する議論はパターン認識の場合と違い, 必ずしも計算論的な意味が十分議論された上で使用されているわけではない．本章ではロボットの位置推定問題における特徴選択のために相互情報量にもとづく評価基準を考察し, PCA 回帰モデルがこの基準を正当化できる条件とその特長を見る.

4.1 KL ダイバージェンスによる評価基準

データセットの発生源となる真の確率分布を仮定すると，これとロボットが場所の推定に用いるパラメトリックモデルとの分布間の距離, KL ダイバージェンス (Kullback-Leibler divergence) を推定の良さの基準として自然に導入することができる．そこで, 真の確率分布を x, y 空間上で定義し, $P^*(x,y)$ とする．これに対して推定に用いるモデル, 式 (4) による確率分布は $P(x,y) = P(y|x;\theta)P(x)$ であるから, この両者の KL ダイバージェンスは, 次のように書ける.

$$\begin{aligned} & \int_x \int_y P^*(x,y) \log \frac{P^*(x,y)}{P(y|x;\theta)P(x)} dy dx \\ = & \int_x \int_y P^*(x,y) \log P^*(x,y) dy dx \\ - & \int_x \int_y P^*(x,y) \log P(y|x;\theta)P(x) dy dx. \end{aligned} \quad (5)$$

第一項は D と $y = f(z)$ により決まる負のエントロピーになる．また第二項はモデルの D に関する対数尤度であり, これを $L_g(\theta)$ と書く.

4.2 相互情報量にもとづく評価基準

KL ダイバージェンス式 (5) の第一項の符号に注意すると, 第二項の尤度が等しい場合にはエントロピーが大きく, 確率分布の分散が大きいモデルをより良いとする基準になっている．これをロボットの位置推定に適用すると, データセット D からの近似が同程度であるとき, 確率モデルによる推定結果の確信度が低い方のモデルを選ぶ基準になる．つまりこの基準によって特徴を選択すると, 写像後の分布 D_f のエントロピーがより大きく, 推定結果の曖昧性を高くしてしまうものが選ばれてしまい不都合が生じる．そこで, ロボットの場所推定問題固有の性質を考慮した評価基準を考える.

推定結果の曖昧性はセンサ入力 z が写像 $y = f(z)$ を通して与えられた時の x の確率分布 $P(x|y = f(z))$ のエントロピー $H(x|y) \equiv -\int_x P(x|y = f(z)) log P(x|y = f(z)) dx$, を全ての y について平均した $E_y[H(x|y)]$ で評価でき, これは次のように計算できる.

$$\int_y \int_x -P^*(x|y) log P^*(x|y) dx P^*(y) dy. \quad (6)$$

ただし,

$$\begin{aligned} P^*(y) & \equiv \int_x P^*(x,y) dx & (7) \\ P^*(x|y) & \equiv P^*(x,y)/P^*(y). & (8) \end{aligned}$$

さらに $E_y[H(x|y)]$ は Bayes の定理から次のように展開できる.

$$E_y[H(x|y)] = E_y[H(y|x) + H(x) - H(y)]. \quad (9)$$

$H(x)$ はデータセットにより決まるので, 写像 $y = f(z)$ を選ぶことで操作できるのは, $H(y|x) - H(y)$ だけである．そこでこれを式 (5) の第一項の代わりに用いた,

$$E_y[H(y|x) - H(y)] - L_g(\theta) \quad (10)$$

を考える．この式 (10) を最小化する y は相互情報量,

$$I(x,y) = H(x) - H(x|y) = -(H(y|x) - H(y)) \quad (11)$$

を最大にする特徴選択であり，直観的に妥当な基準である.

この相互情報量を実際に計算する場合は，写像後のデータからヒストグラムを得てエントロピーを計算することができる．またデータがパラメトリックモデルによって十分近似できる場合にはこれを解析的に求めることもできる．そこで, PCA 回帰モデルの場合にこれを計算してみると, $H(y|x)$ は式 (4) から, 分散 σ のみに依存する量となる (σ は回帰で仮定する定数)．また, PCA により選ばれた m 番目の特徴 $y^{(m)}$ は分散を最大になるような順に選ばれているから $m < n$ においては $H(y^{(m)}) > H(y^{(n)})$ が成立している．したがって PCA

回帰モデルの特徴選択は式 (10) の第一項を最小化することが言える.

次に第二項について見ると, 式 (10) の積分には, 未知の真の分布 $P^*(x,y)$ が必要となるので, その代用として離散的に得られている経験データ D による総和から計算する. すると PCA 回帰モデルの場合の式 (10) の第二項は, 式 (4) を代入して次のようになる.

$$\begin{aligned} L_g(\theta) &\propto \sum_i \log P(y=f(z_i)|x_i;\theta) \\ &= \sum_i \log \frac{exp(-(g(x_i;\theta)-f(z_i))^2/2\sigma^2)}{\sigma\sqrt{2\pi}} \\ &= \alpha - \beta \sum_i (g(x_i;\theta)-f(z_i))^2. \end{aligned} \quad (12)$$

ただし, α, β は σ から定まる定数である.

したがって, これを最適化するための θ は二乗誤差 $\sum_i (g(x_i;\theta)-f(z_i))^2$ を最小化するように選べば良く, 最小二乗基準により決定した PCA 回帰モデルは式 (10) の第二項を最適化することが言える.

以上の結果から, PCA 回帰モデルは後段の回帰モデルが十分小さい分散 σ の仮定のもとで当てはまるならば, 我々の提案するエントロピーにもとづく基準 (10) を最適化するものであることがわかる.

4.3 Avg. Bayesian localization error

S.Thrun[4] は次に紹介する Averaged Bayesian Localization Error とこれを最適化するようにニューラルネットを学習することで得られる非線形な特徴選択法を提案している.

まずロボットが位置している真の場所を x^*, 推定場所を x とした時の推定誤差はノルム $||x-x^*||$ で評価できる. そこで, 真の x^* を一つ決めた時に確率モデル $P(x|y)$ の平均推定誤差を次のように表す.

$$Err(x^*) = \int_x \int_y ||x-x^*|| P(y|x^*) P(x|y) dx. \quad (13)$$

ここで, $P(x|y)$ は入力が y であった時の推定場所 x の確率分布, $P(y|x^*)$ は真の場所 x^* のもとでの入力 y の尤度であり, 上式は全ての取り得る y の値について, 真の場所が x^* であった時の確率モデルの推定誤差の期待値を意味している. 先に述べたように, 確率モデルは $P(y|x)$ の形で定義されるので上式はベイズの定理を用いて

$$\begin{aligned} Err(x^*) = \int_x \int_y ||x-x^*|| P(y|x^*) \\ P(y|x)P(x)P^{-1}(y)dydx, \end{aligned} \quad (14)$$

$$P^{-1}(x) = \int_x P(y|x)P(x)dx. \quad (15)$$

となる. さらにこれをロボットが存在し得る全ての真の場所 x^* の事前分布 $P(x^*)$ による期待値をとることで, 最終的に

$$\begin{aligned} Err &= avg_{x^*} Err(x^*) \\ &= \int_x^* Err(x^*)P(x^*)dx^* \\ &= \int_{x^*}\int_x\int_y ||x^*-x|| P(y|x^*) \\ &\cdot \; P(y|x)P(x)P(x^*)P^{-1}(y)dydxdx^*. \end{aligned} \quad (16)$$

が得られる. これが Avaraged Bayesian Localization Error である. 文献 [4] では y を真／偽二値のバイナリ変数に限定し, これをある種の非明示的なランドマークと考え, この基準を最適化するようにニューラルネットを最急降下的に学習することで自動的なランドマーク決定と特徴選択を行っている.

この Averaged Bayesian localization error の計算は x^*, x, y の三変数についての確率平均をとるため, 計算量は $\mathcal{O}(|y||x|^2)$ になる. ($|x|$ は x の確率空間のサイズ). さらにニューラルネットの学習のために勾配の計算, 学習回数を考慮しなければならないので, 特徴選択に要する計算コストがかなり大きい.

一方, 本稿で述べた評価基準式 (10) は x, y の 2 変数についての周辺化であるから, 計算量は $\mathcal{O}(|y||x|)$ となりより低コストであり, PCA 回帰モデルのパラメータ決定は繰り返し的な計算を必要としないというメリットがある.

5 Conditional mixture of PCA

4.2 で述べたように, 相互情報量にもとづく評価基準を最適化するためには y, x 間の確率モデルの分散を小さくすることが必要である. そして理想的な特徴選択が得られた場合には y, x 間の回帰は相関の高いものになる (図 2).

しかし, z, x 間の非線形性が強い場合には, これを z, y 間の写像 $y=f(z)$ で解消しなければならない. しかし, 従来の PCA は $f(z)$ を大域的な線形写像として構成するものであり, 後段に非線形性が残ってしまう場合がある (図 3).

そこで PCA の非線形化を考える. 最近局所的な PCA による写像を結合する Mixture of PCA による非線形化が考案されている [1, 2]. これは, 複数の異なる PCA の写像 $f_i(z)$ を用意して次のように書ける.

$$f_{mix}(z) = \sum_i \pi_i(z) f_i(z), \quad (17)$$

ここで $\pi_i(z)$ は重み係数で $\sum \pi_i = 1$, $\pi > 0$ である. こ

図 2: 理想的な特徴選択 (x 軸: ロボットの X 座標, y 軸: 全方位カメラ画像から PCA で選択した第一主成分)

図 3: 良い推定が得られない場合 (x 軸: ロボットの X 座標, y 軸:全方位カメラ画像から PCA で選択した第一主成分)

図 4: この場合, 領域 A ではソナー 3,7 が, 領域 B では 1,5 が高い寄与率を与える.

の $\pi_i(z)$ の決定には EM アルゴリズムが用いられることが多い [1, 2].

このMixture of PCA は z, y 間で非線形写像を実現するが, x に関しては大域的に単一の写像 $f(z)$ を構成するものである. Robot localization の場合には適用できる環境の領域を拡大できることが重要である. PCA で x がある特定の領域内に限定されている時にはとても寄与率の高い特徴ベクトルを選べたとしても, これが異なる別の領域でも高い寄与率を与えるとは限らない (例: 図 4). このような場合には対象となる全ての領域について大域的に決定した特徴は, 各局所領域において個別に決定した特徴よりも高い場所推定精度を与えることができない. 特にロボットが移動する環境が廊下や部屋など, 特徴的なモジュールの組合せからなり, 全体の領域が広大になるにつれて, このようなモジュールを多数含むことが多く, こうした問題が顕著になる.

したがって実際的な応用において, 適用できる領域の拡大に対応するためには, 該当する局所領域をセンサデータ z 以外の粗い情報から推定し, その局所領域に対して特に良い特徴選択を与える写像に切り替えるような階層的なモデルが必要となる.

自律 Robot の移動の場合には, 過去の位置 x^{t-1} とそこでとったアクション a^{t-1} から次の位置 x^t を条件付き確率 $P(x^t|x^{t-1}, a^{t-1})$ を表すベイジアンネット [11] によってモデル化できる (図 5). そこで式 (17) の係数 $\pi_i(z)$ の代わりに条件付き確率を用いて次のようなモデル, Conditional Mixture of PCA を考える.

$$f_{cond}(z) = \sum_i P(x^t \in R_i|x^{t-1}, a^{t-1}) f_i(z), \qquad (18)$$

ここで R_i は i 番目の写像が対応する領域である.

このモデルは人間が日常的に自然に行っているような

特定の文脈や状況によって起こりやすい仮説を念頭におき, それを立証するためにもっとも顕著な特徴に注意を向けるといった認知的行動ともよく合致している.

図 5: ベイジアンネットでの条件付き確率のモデル化

このような階層的なモデルを用いた精緻化によって, 本稿で述べた評価基準の意味でより良い特徴選択を実現できると期待できる.

6 おわりに

本稿では, Robot localization タスクにおける特徴選択に焦点をあて, 確率的なモデル化と相互情報量に基づく評価基準について議論を行った. これらを通じて, PCA 回帰モデルの計算論的意味と PCA による特徴選択が良い場所推定を与える条件を示した. しかしロボットが移動できる領域を拡大するために PCA が大域的に単一の線形写像であることが問題となる場合があり, このための非線形化と, 条件付き確率と組み合わせた混合モデル, Conditional mixture of PCA を用いる方法についても述べた.

謝辞

有益な議論をいただいた, Roland Bunschoten, Bert Kappen, Wim Wiegerinck, David Barber の各氏に感謝する. また本研究は Real World Computing プログラムの一貫として行われ, RWCP より実環境ロボットデータベース MEMORABLE の提供を受けた.

参考文献

[1] M.Tipping and C.Bishop, “Mixture of Probabilistic Principal Component Analysis”, *Proc. of the ICANN’97*, 1997.

[2] G.Hinton, M.Revow and P.Dayan, “Recognizing handwritten digits using mixture of linear models”, *Advances in Neural Information Processing 7*, 1995.

[3] T.Dean and M.Wellman, “Planing and control”, Morgan Kaufmann,CA ,1991.

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

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

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

[7] 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.

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

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

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

[11] 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.

[12] N.Vlassis, Y.Motomura, and B.Krose, “An Information-theoretic localization criterion for robot map building”, *ACAI ’99*, 1999, *to apear.*

[13] Y.Motomura, N.Vlassis and B.Krose, “Probabilistic Robot Localization and Situated Feature Focusing”, *IEEE SMC’99*, 1999, *to apear.*