1O3-3

語彙と指差し位置の同時推定による場所概念獲得

Location Concept Acquisition Based on Simultaneous Estimation of Word and Pointing Position

小林博慶 Hiroyoshi Kobayashi　谷口彰 Akira Taniguchi　萩原良信 Yoshinobu Hagiwara　谷口忠大 Tadahiro Taniguchi

立命館大学
Ritsumeikan University

Human support robots that move in human living environments, need to learn locations by communicating with people. When a person teaches a location to a person, pointing motions are used to indicate the location. In this paper, we propose a robot system to learn the name of place by estimating pointing motions by a person. It is expected that the system enables robots to learn locations in a remote places. We did an experiment that acquire the concept of location using the two methods of pointing. The first method, the user teach the location without moving a finger. The second method, the user teach the location by moving a finger so as to surround the region of the location. The robot was able to acquire concept of location where the user is pointing.

1. はじめに

人間の生活環境で動作するロボットは，人とのコミュニケーションをとるために場所の概念について学習する必要がある．例えば，ロボットが人に「キッチンにあるポットを取ってきて」と頼まれたとき，ロボットがキッチンという場所の概念を持っていなければ，どこにキッチンがあるのか分からずポットを取ってくることができない．本研究において場所の概念とは，場所の名前とその名前と対応した位置分布によって表されるものとする．本研究では，ある特定の地点を位置と定義し，位置の空間的な広がりを位置分布と定義する．

場所の概念を学習する研究として谷口らの研究がある [1]．谷口らは，不確実な音声認識結果と自己位置推定結果を統合して場所の概念をロボットに獲得させることを実現している．この研究では，人の教示に基づいてロボットが場所の概念について学習する．ただし，場所の概念を学習するとき，教示の対象となる場所にロボットが移動する必要がある．しかし，場所の概念についてひとつひとつ順番に移動しながら教えることは時間的なコストがかかるという課題がある．また，人が移動できないような場所について教えることができないという課題がある．人が場所を教えるときには，わざわざその場所までいかなくとも，指でその場所を指し示すことによって教えることができる．本研究の目的は，ロボットが指差しを用いて場所の概念を学習できるようにすることでこれらの課題を解決することである．

本研究では，人の指差しを用いた教示によってロボットが場所の概念の学習を効率的に行うモデルを提案する．本稿では，指を動かさずに場所を教示する方法と場所の空間的な広がりを囲うように指を動かしながら場所を教示する方法の 2 つの指差し方法を用いた場所の概念獲得の実験を行い，提案手法を用いて場所の概念獲得が可能であるかを検証する．

2. 指差しに基づく場所の概念の学習手法

提案手法では，人が教示対象の場所を指差しながらその場所の名前の教示を行う．これを複数回繰り返し，人に教示された単語と指差し位置から場所の名前と場所の位置分布を同時に推定して場所の概念を学習する．

本研究では，ロボットの姿勢をパーティクルで表現する自己位置推定の手法である MCL(Monte Carlo Localization)[2][3] に場所の概念を導入した谷口らのモデル [1] に，指差し位置の推定を導入したモデルを提案する．

提案手法を用いることでロボットは自身の自己位置から離れた場所の概念について学習することができるようになり，場所の概念の学習を効率的に行うことができる．

連絡先: 小林博慶，立命館大学，滋賀県草津市野路東 1 － 1 － 1，TEL 077-561-5745，FAX 077-561-5745，h.kobayashi@em.ci.ritsumei.ac.jp

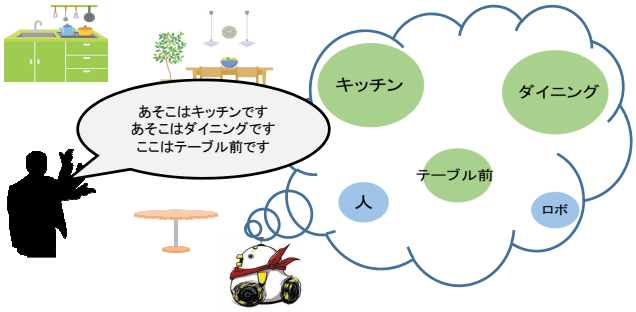


図 1: 想定するタスクの概要図

3. 想定するタスクの概要

本研究では，ロボットが人の指差しを認識し，指差し位置の推定を行うことを想定する．例としてキッチン，ダイニング，テーブル前の三つの場所を学習対象の場所とすることを考える．例えば，図 1 のように，学習対象の場所から離れた位置に人とロボットがいるとき，人がキッチンを指差しながらロボットに「あそこがキッチンです」と教示する．次に，人もロボットも位置を移動 せずに人がダイニングを指差しながらロボットに「あそこがダイニングです」と教示する．テーブル前も同様に行う．図 1 のように，指差しした場所付近に位置分布が構成され，その分布に対応した場所の名前が学習される．

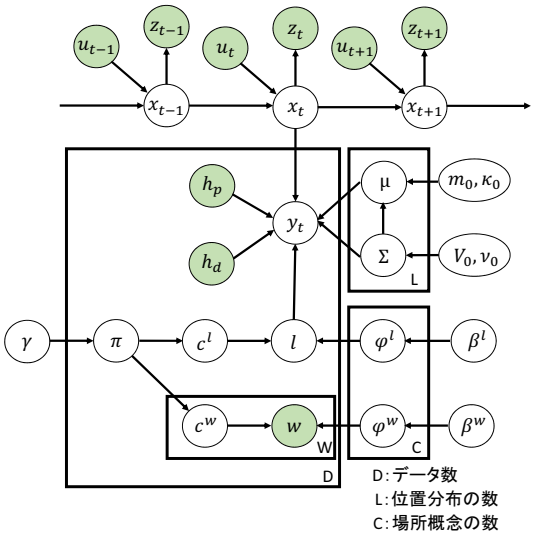


図 2: 提案手法のグラフィカルモデル

表 1: 提案手法のグラフィカルモデルの要素

x_t	ロボットの自己位置
u_t	制御値
z_t	計測値
y_t	指差しされた位置
h_p	ロボットからみた人の相対位置
h_d	ロボットからみた人の指差し方向
l	位置分布の index
μ, Σ	位置分布 (平均，共分散)
π	カテゴリの多項分布
c^l, c^w	カテゴリ
w	単語 (Bag-of-Words)
φ^l	位置分布の index の多項分布
φ^w	単語の多項分布
m_0, κ_0, V_0, v_0 β^l, β^w, γ	事前分布の ハイパーパラメータ

4. 提案モデル

提案手法のグラフィカルモデルを図 2 に示す．提案手法のグラフィカルモデルの各パラメータについての説明を表 1 に示す．このモデルでは，指差しと発話による教示からそれぞれの場所を表す位置分布と場所の名前を学習する．ロボットは事前に単語知識を持っており，誤りなく単語を認識するものとする．

4.1 指差し位置推定手法

人の指差しを認識するために Xbox One Kinect センサー (Kinect v2) [*1] を用いる．Kinect v2 ではセンサーに写っている人の骨格データを 25 カ所の関節の位置の 3 次元座標データで取得することができる．本研究では，右手を用いて指差しによる場所の教示を行うことを想定する．Kinect v2 で取得した頭と右手先の座標を直線で結び地面との交点を一時的な指差し位置座標とする．本研究では，人の指差しの位置には誤差が生じているものとする．そのため，この一時的な指差し位置座標を平均，指差し位置の誤差を分散としたガウス分布によって，Kinect v2 によって推定された指差し位置の信頼度合いやあいまいさを表現することとする．一時的な人が指差しした位置は (13) 式のようなガウス分布に従うと定義する．

4.2 生成モデル

本提案手法の生成モデルを (1 – 12) 式のように定義する．

$$
\begin{aligned}
\pi &\sim \mathrm{Dir}(\gamma) && (1)\\
C^{\beta} &\sim \mathrm{Mult}(\pi) && (2)\\
C^{w} &\sim \mathrm{Mult}(\pi) && (3)\\
\varphi^{l} &\sim \mathrm{Dir}(\beta^{l}) && (4)\\
\varphi^{w} &\sim \mathrm{Dir}(\beta^{w}) && (5)\\
\mu &\sim \mathcal{N}(\mu|m_0, (\kappa_o\Lambda)^{-1}) && (6)\\
\Sigma^{-1} &\sim \mathcal{W}(\Sigma^{-1}|V_0, v_0) && (7)\\
l &\sim \mathrm{Mult}(\phi_{C^l}) && (8)\\
w &\sim \mathrm{Mult}(\phi_{C^w}) && (9)\\
y_t &\sim p(y_t|l, \mu, \Sigma, x_t, h_p, h_d) && (10)\\
x_t &\sim p(x_t|x_{t-1}, u_t) && (11)\\
z_t &\sim p(z_t|x_t) && (12)
\end{aligned}
$$

ここで，(10) 式は (13) 式のように定義する．

$$
\begin{aligned}
& p(y_t|l, \mu, \Sigma, x_t, h_p, h_d)\\
=\ & p(y_t|\mu_l, \Sigma_l, x_t, h_p, h_d)\\
\propto\ & \mathcal{N}(y_t|\mu_l, \Sigma_l) p(y_t|x_t, h_p, h_d)\\
=\ & \mathcal{N}(y_t|\mu_l, \Sigma_l)\mathcal{N}(y_t|y', \sigma^2) \qquad (13)
\end{aligned}
$$

ここで，y' は kinect で推定された指差し位置の平均，σ^2 は kinect で推定された指差し位置の分散を表している．

4.3 場所概念学習手法

学習は，複数回教示されたデータを溜め込み，オフラインで行う．このとき教示された時刻 t の集合を $T = \{t_1, t_2, \cdots, t_D\}$ とする．D は教示データ数である．時刻ごとの制御値，計測値，ロボットからみた人の相対位置，ロボットからみた人の指差し方向および人に教示された単語による複数の教示データから，モデルのパラメータを周辺化ギブスサンプリングによって推定する．

教示の際は，自己位置推定するロボットの正面に人が立つ．人は人とロボットから離れた位置にある教示対象の場所を指差しながら場所の名前の入力を行う．

位置分布の初期値は全て μ_l =(一定の範囲内に一様乱数)，$\Sigma_l = \begin{bmatrix} \sigma_{initial} & 0 \\ 0 & \sigma_{initial} \end{bmatrix}$ とする．

以下に，周辺化ギブスサンプリングを行う際の各パラメータごとの事後分布を表す．

(14) 式は，位置分布の index l に関する事後分布である．

$$
\begin{aligned}
l \ &\sim\ p(l|C^l, \beta^l, y_t, h_p, h_d, \mu, \Sigma, x_t)\\
&\propto\ \frac{n_{c,l}^{\backslash d,i} + \beta_l}{n_{l,\cdot}^{\backslash d,i} + \Sigma_{l'}\beta_{l'}} \mathcal{N}(y_t|\mu_l, \Sigma_l)\mathcal{N}(y_t|y', \sigma^2) \qquad (14)
\end{aligned}
$$

*1 http://www.xbox.com/ja-JP/xbox-one/accessories/kinect-for-xbox-one

(15) 式は，カテゴリ c^l に関する事後分布である．

$$\begin{aligned} c^l &\sim p(c^l_{d,i} = c | l_{d,i} = l, l^{\setminus d,i}, c^{l\setminus d,i}, \gamma, \beta^l) \\ &\propto \int p(l_{d,i} = l | \phi^l_c) p(\phi^l_c | l^{\setminus d,i}, c^{l\setminus d,i}, \beta^l) d\phi_c \cdot \\ &\quad \int p(c^l_{d,i} = c | \pi_d) p(\pi_d | c^{l\setminus d,i}, \gamma) d\pi \\ &= \frac{n^{\setminus d,i}_{c,l} + \beta_l}{n^{\setminus d,i}_{l,\cdot} + \Sigma_{l'} \beta_{l'}} \frac{n^{d,i}_{d,c} + \gamma_c}{n^{\setminus d,i}_{d} + \Sigma_{c'} \gamma_{c'}} \end{aligned} \tag{15}$$

ここで，$n^{\setminus d,i}_{c,l}$ は全データに対して潜在トピック c が位置分布の index l に対して推定された回数である．$n^{d,i}_{d,c}$ はデータ d で潜在トピック c が現れた回数である．$n^{\setminus d,i}_{l,\cdot}$ は潜在トピック c の総割り当て数である．$n^{\setminus d,i}_{d}$ はデータ d の位置分布の index の観測回数である．

c_l と同様に式 (16) で c_w が求められる．

$$\begin{aligned} c^w &\sim p(c^w_{d,i} = c | w_{d,i} = w, w^{\setminus d,i}, c^{w\setminus d,i}, \gamma, \beta^w) \\ &= \frac{n^{\setminus d,i}_{c,w} + \beta_w}{n^{\setminus d,i}_{w,\cdot} + \Sigma_{w'} \beta_{w'}} \frac{n^{d,i}_{d,c} + \gamma_c}{n^{\setminus d,i}_{d} + \Sigma_{c'} \gamma_{c'}} \end{aligned} \tag{16}$$

式 (17) は指差しされた位置 y_t に関する事後分布である．

$$\begin{aligned} y_t &\sim p(y_t | l, \mu, \Sigma, x_t, h_p, h_d) \\ &\propto \mathcal{N}(y_t | \mu_l, \Sigma_l) \mathcal{N}(y_t | y', \sigma^2) \end{aligned} \tag{17}$$

位置分布 μ, Σ は，$l \in L$ ごとにサンプリングできる．このとき，$m_{n_k}, \kappa_{n_k}, V_{n_k}, \nu_{n_k}$ は事後パラメータであり，y_k は $t \in T$ の中で $l_t = k$ である教示位置を集めたものである．

$$\begin{aligned} \mu, \Sigma &\sim \mathcal{N} - \mathcal{W}(\mu_k, \Sigma_k | m_{n_k}, \kappa_{n_k}, V_{n_k}, \nu_{n_k}) \\ &\propto \mathcal{N}(x_k | \mu_k, \Sigma_k) \\ &\times \mathcal{N} - \mathcal{W}(\mu_k, \Sigma_k | m_0, \kappa_0, V_0, \nu_0) \end{aligned} \tag{18}$$

ロボットの自己位置のサンプリングに関しては，時刻 t に対する教示の有無で分ける．

$$\begin{aligned} x_t &\sim p(x_t | x_{t-1}, x_{t+1}, u_t, u_{t+1}, z_t) \\ &\propto p(x_{t+1} | x_t, u_{t+1}) p(z_t | x_t) p(x_t | x_{t-1}, u_t) \end{aligned} \tag{19}$$

式 (19) は時刻 t に教示を行っていない場合の自己位置 x_t の事後分布である．

$$\begin{aligned} x_t &\sim p(x_t | x_{t-1}, x_{t+1}, u_t, u_{t+1}, z_t, y, h_p, h_d, l, \mu, \Sigma) \\ &\propto p(x_{t+1} | x_t, u_{t+1}) p(z_t | x_t) p(x_t | x_{t-1}, u_t) \\ &\times \mathcal{N}(y_t | \mu_l, \Sigma_l) \mathcal{N}(y_t | y', \sigma^2) \end{aligned} \tag{20}$$

式 (20) は時刻 t に教示を行った場合の自己位置 x_t の事後分布である．

5. 場所概念学習実験

5.1 実験目的

本実験では，提案モデルで学習した場所の概念が 2 つの指差し方法の違いによってどのように変化するかを確認する．

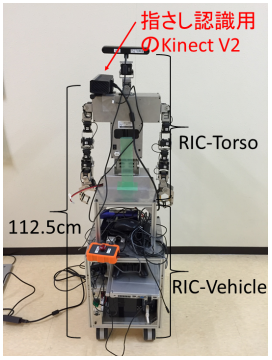


図 3: RIC Torso と RIC Vehicle

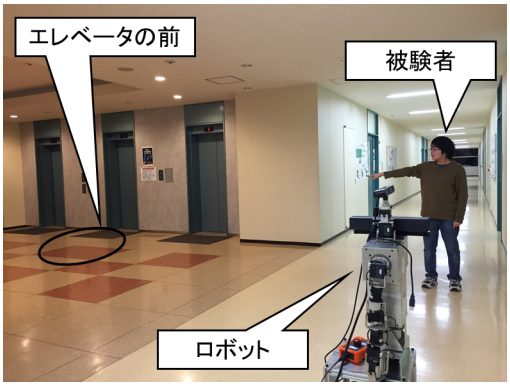


図 4: 実験環境と実験の様子

5.2 実験条件

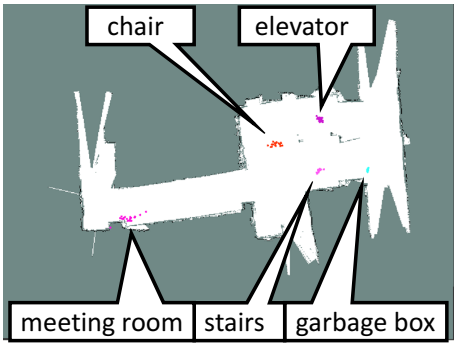
ロボットは RT ROBOTSHOP 社製の RIC-Torso 及び RIC-Vehicle(図 3[*2]) を用いる．ロボットと人は決められた位置に立ち移動しない．ロボットは Kinect v2 を搭載し，人の指差しを認識する．場所概念の数を $C = 5$，位置分布の数を $L = 5$ と設定する．各パラメータはそれぞれ $\gamma = 10$，$\beta^c = 1$，$\beta^w = 1$，$\kappa_0 = 1e-3$，$V_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$，$v_0 = 2$ とする．イテレーション回数は 100 回である．ロボットの自己位置 x_t については推定せず，既知とする．学習対象の場所はエレベータの前 (elevator)，ゴミ箱の前 (garvage box)，階段の前 (stairs)，椅子の前 (chair)，会議室の前 (meeting room) の 5 カ所とする．名前の教示は人が発話して教示した結果を誤りなく認識したと仮定してキーボードから手入力で行う．指差しによる場所の位置の教示は 2 つの方法に分けて行う．方法 1 は各場所に対して指を指し続けたまま 20 回ずつ計 100 回名前を教示する．方法 2 は各場所に対してその場所に対応する位置分布全体を囲むように指を動かしながら 20 回ずつ計 100 回名前を教示する．

図 4 に実際の実験環境と実験の様子を示す．エレベータホールのエレベータの前とロボット，教示している人を示している．図は人がエレベータの前を教示している際の様子である．

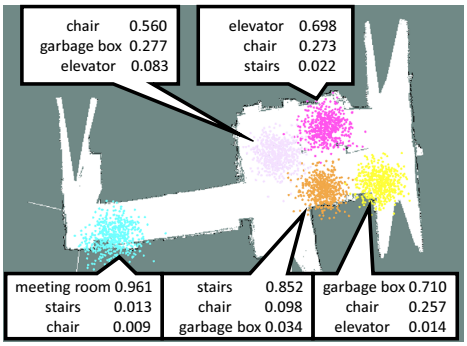
5.3 実験結果

図中の地図はロボットがオドメトリ情報と距離センサを用いて姿勢と地図の同時推定 (SLAM) を行いながら作成したエレベータホール前の地図である．図 5(b) 図 6(b) の点群は学習

*2 http://products.rt-net.jp/ric/ric-torso

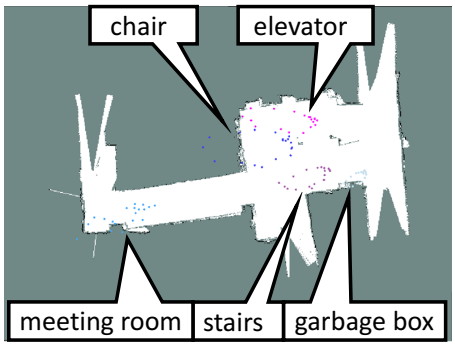


(a) 教示した位置データ

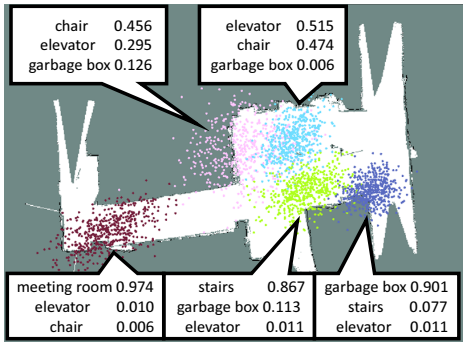


(b) 場所の概念の学習結果

図 5: 方法 1 の指差しによる実験の結果



(a) 教示した位置データ



(b) 場所の概念の学習結果

図 6: 方法 2 の指差しによる実験の結果

した場所の概念の位置分布に従う点を 500 個ずつ描画したものである．場所の位置分布はガウス分布で表されるので，学習した位置分布の平均と分散を用いたガウス分布からランダムにサンプリングした結果を描画している．異なる位置分布 (ガウス分布) のサンプリング結果はそれぞれ別の色で表現した．それぞれのふきだしは各位置分布に対応した場所のインデックスである．図 5(a) 図 6(a) の点群は Kinect v2 を用いて推定した指差し位置のガウス分布の平均 y' を教示 100 回分描画したものである．異なる教示場所の点を異なる色で表現した．

この結果から，方法 1 ではそれぞれの位置分布に対応する場所の名前として，正しいと考えられる単語の確率が高いことが分かった．方法 2 では 5ヶ所中 3ヶ所は場所の名前として正しいと考えられる単語の確率が高かったが，2 ヶ所について間違った単語の確率が高くなっていた．方法 1 では最も各位置分布の広がりが小さくなり，どの位置分布も同じような広がりとなった．方法 2 が最も各位置分布の広がりが大きく，位置分布同士が重なっている部分が見られた．

6. おわりに

本稿では，場所の概念を効率的に学習することを目的として，谷口らの提案した場所の概念の学習モデルに人の指差しを認識して推定する機能を追加することで場所の名前と指差し位置を同時推定しながら人やロボットから離れた位置にある場所の概念を学習するモデルを提案した．1 点を指差し続けながら 20 回教示する方法では，場所ごとの位置分布の広がりの違いを表現できていなかったが位置分布と場所の名前は正しく対応して表現できていた．各場所の位置分布を囲うように指を動かしながら教示する方法では位置分布と場所の名前の対応が正しくできていない場合があったものの，概ね正しく対応しており，場所ごとの空間的な広がりの違いを表現することができた．これらの結果から，提案手法を用いて指差しによる場所の概念の獲得が可能であることを確認した．

今後は，例えば位置分布が同じで名前が複数ある場所や名前が同じで複数の位置分布がある場所のような複雑な実験環境での検証を行いたい．今後の展望として，場所の名前を人の発話文から推定できるようにすることや画像情報と統合すること [4] でより人のもつ場所の概念に近い場所の概念の獲得を可能にすることが挙げられる．

参考文献

[1] Akira Taniguchi, Tadahiro Taniguchi, and Tetsunari Inamura. Spatial concept acquisition for a mobile robot that integrates self-localization and unsupervised word discovery from spoken sentences. *arXiv preprint*, Vol. airXiv:1602.01208, , 2016.

[2] Thrun Sebastian, Burgard Wolfram, and Fox Dieter. 確率ロボティクス. 毎日コミュニケーションズ, 2007.

[3] 友納正裕. 移動ロボットのための確率的な自己位置推定と地図構築. 日本ロボット学会誌, Vol. 29, No. 5, pp. 423–426, 2011.

[4] Satoshi Ishibushi, Akira Taniguchi, Toshiaki Takano, Yoshinobu Hagiwara, and Tadahiro Taniguchi. Statistical localization exploiting convolutional neural network for an autonomous vehicle. In *Industrial Electronics Society, IECON 2015-41st Annual Conference of the IEEE*, pp. 001369–001375. IEEE, 2015.