1I4-OS-09a-3

発話文の教師なし形態素解析と位置推定を統合したノンパラメトリックベイズ場所概念獲得

Nonparametric Bayesian Location Concept Acquisition that Integrates Localization and Unsupervised Word Segmentation of Utterance Sentence

谷口 彰[*1] 稲邑 哲也[*2*3] 谷口 忠大[*1]
Akira Taniguchi Tetsunari Inamura Tadahiro Taniguchi

[*1]立命館大学 Ritsumeikan University
[*2]国立情報学研究所 National Institute of Informatics
[*3]総合研究大学院大学 The Graduate University for Advanced Studies

In this paper, we propose a novel learning method which can estimate self-location of a robot and concepts of location simultaneously. A robot performs a probabilistic self-localization from sensor data. We propose nonparametric bayesian location concept acquisition that integrates localization and unsupervised word segmentation of utterance sentence.

1. はじめに

人間の生活環境下で動作するロボットは，様々な環境において周囲の様子を認知し，人間とのインタラクションを通して環境中の場所に対し人間が割り当てた語彙と，その語が指し示す空間領域を学習することが重要である．このとき，センサのノイズ，移動誤差，音声認識誤りなどの多くの不確実性への対処が重要となる．本研究では，事前に語彙を持たず日本語音節のみを認識可能で，自己位置推定を行いながら環境を移動するロボットに，人間が場所の名前を発話文により教示することで，場所に対応した語彙を獲得させることを目的とする．

以上の目的の下，我々は不確実な音声認識結果と自己位置推定情報を相互に有効活用した，自己位置と語彙の同時推定モデルを提案している [1]．本稿では，一単語発話しか学習できなかった上記のモデルを複数単語文扱えるように拡張した，発話文の教師なし形態素解析と位置推定を統合したノンパラメトリックベイズ法による場所概念獲得モデルを提案する．

2. 先行研究

語彙を持たないロボットに，多様な言い回し発話から単語の正しい分節，音素系列，単語と対象間の対応関係を学習させる手法が提案されている [2]．山田らの研究では，先の手法 [2] を拡張し，自己位置座標のカテゴリ化と語彙学習を同時に行う手法が提案されている [3]．しかし，学習した言語知識をロボット自身の自己位置推定タスクに有効活用することはできていない．本研究では，音節認識誤りのある多様な言い回しの発話文から場所に関する語彙獲得を行い，さらにそれを自己位置推定に有効活用する手法を提案する．

3. 自己位置と語彙の推定モデル

本研究では，環境中のある特定の座標や局所的な地点のことを位置と呼び，位置の空間的な広がりを位置分布とする．場所概念とは，場所の名前とその名前と対応したいくつかの位置分布によって表されるものとする．本研究では，状態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 Localizatoin)[4] に場所概念を導入したモデルを提案する．本研究では主として，(1) 音節認識誤りあり発話文からの単語の分節化と，(2) 場所の名前を複数回教示されたときの場所概念の学習方法，(3) 場所概念を獲得したロボットが場所の名前を聞いたときの自己位置推定について問題とする．

連絡先: 谷口彰，立命館大学情報理工学研究科，a.taniguchi@em.ci.ritsumei.ac.jp

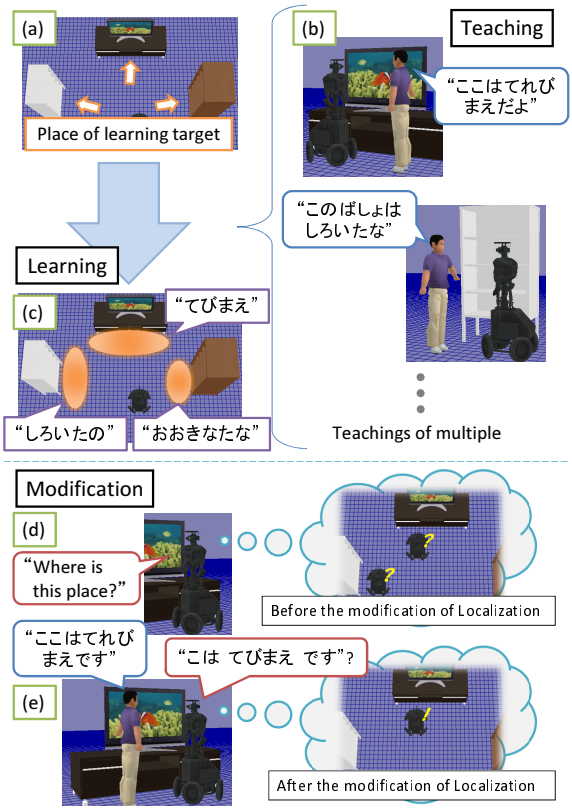

図 1: Schematic diagram of the proposed method

3.1 提案モデルとタスクの概要

事前に環境の地図を持った移動ロボットを動作させ，自己位置推定を行わせることを想定する．提案手法の全体像を表す概略図を図 1 に示す．図 1 (a) の様に，三つの各物体前の場所付近を学習対象の場所とする．例えば，図 1 (b) の様に，人とロボットがテレビの前にいるとき，人がロボットに “ここはてれびまえだよ” と発話し教示を行う．白い棚付近に移動したと

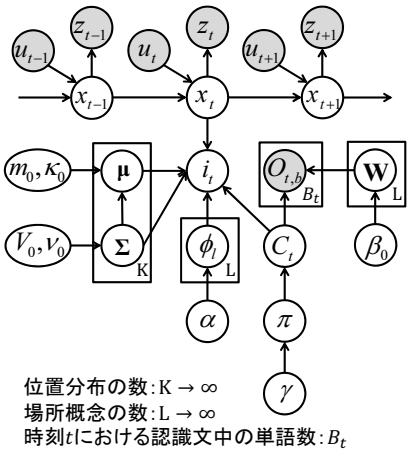

図 2: Graphical model of the new proposed method

表 1: Each element of the graphical model

x_t	ロボットの自己位置
u_t	制御値
z_t	計測値
C_t	場所概念の index
$O_{t,b}$	b 番目の音声認識単語
$\mathbf{W}$	場所の名前（多項分布）
$\boldsymbol{\mu}, \boldsymbol{\Sigma}$	位置分布 (平均，共分散行列)
i_t	位置分布の index
ϕ_l	位置分布の index の多項分布
π	場所概念の index の多項分布
α	ϕ_l のハイパーパラメータ
γ	π のハイパーパラメータ
β_0	ディリクレ事前分布のハイパーパラメータ
$m_0, \kappa_0,$ V_0, ν_0	ガウス-ウィシャート事前分布の ハイパーパラメータ

きは，“このばしょはしろいたな” と教示する．大きな棚の前でも同様である．教示を複数回行った後，ロボットは聞きとった言葉を形態素解析し，場所概念の学習を行う．図 1 (c) の様に，教示した各物体前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．この場合，ロボットは音声認識誤りを含むため，“しろいたの” のような，誤りを含んだ名前が学習される場合も考えられる．学習後，ロボットは自己位置推定を行いながら移動している．図 1 (d) の様に，ロボットは実際にはテレビの前にいるが，自己位置推定の結果はテレビの前か白い棚付近となっている．このときロボットが，人に “ここはどこか？” と尋ねたとする．図 1 (e) の様に，人は “ここはてれびまえです” と発話する．するとロボットは，発話された場所の名前と学習した場所概念を利用して，自身がテレビの前にいる確率が高いことを知り，自己位置推定の情報を修正することができる．

3.2 旧モデル [1] からの拡張点

MCL に場所概念を導入した新たなモデルのグラフィカルモデルを図 2 に示す．グラフィカルモデルの各要素についてまとめたものを表 1 に示す．

これまでのモデルの課題としては，学習の際，場所概念の数を既知としていたことや，一単語発話のみからしか学習できなかったことがあった．これに対し新たなモデルでは，ノンパラメトリックベイズ拡張することによりデータに応じて適切な場所概念の数を学習できるようになる．具体的には，Dirichlet Process の構成法の一つである SBP(Stick Breaking Process)[5] を用いる．また，発話文からの学習も可能になる．発話文に関しては，G. Neubig らの連続音声認識による単語ラティスから教師なし形態素解析を行う手法 [6] を用いて，単語分割と言語モデルの学習を事前に行う．これにより，発話認識結果のゆらぎを抑えることができる．

場所概念について，これまでのモデルでは，場所の名前 (単語) に対して位置分布 (ガウス分布) が一対一対応であった．これに対し，新モデルにおける場所概念は，場所の名前 W_l とそれに対応する多項分布 ϕ_l が示す位置分布 (μ_k, Σ_k) で表される．つまり，場所の名前 (多項分布) の一つに対し複数の位置分布 (混合ガウス分布) が対応可能となる．

3.3 生成モデル

本提案手法の生成モデルを (1-10) 式の様に定義する．

$$\pi \sim \mathrm{GEM}(\gamma) \quad (1)$$
$$C_t \sim \mathrm{Mult}(\pi) \quad (2)$$
$$W \sim \mathrm{Dir}(\beta_0) \quad (3)$$
$$O_{t,b} \sim \mathrm{Mult}(W_{C_t}) \quad (4)$$
$$\phi_l \sim \mathrm{GEM}(\alpha) \quad (5)$$
$$i_t \sim p(i_t \mid x_t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, C_t) \quad (6)$$
$$\Sigma^{-1} \sim \mathcal{W}(\Lambda \mid V_0, \nu_0) \quad (7)$$
$$\mu \sim \mathcal{N}(\mu \mid m_0, (\kappa_0 \Lambda)^{-1}) \quad (8)$$
$$x_t \sim p(x_t \mid x_{t-1}, u_t) \quad (9)$$
$$z_t \sim p(z_t \mid x_t) \quad (10)$$

ここで，(6) 式は，(11) 式の様に定義する．

$$p(i_t \mid x_t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, C_t) = \frac{\mathcal{N}(x_t \mid \mu_{i_t}, \Sigma_{i_t})\mathrm{Mult}(i_t \mid \phi_{C_t})}{\sum_{i_t=j} \mathcal{N}(x_t \mid \mu_j, \Sigma_j)\mathrm{Mult}(j \mid \phi_{C_t})} \quad (11)$$

$p(x_t \mid x_{t-1}, u_t)$，$p(z_t \mid x_t)$ は，MCL の動作モデル，計測モデルである．

3.4 場所概念の学習

複数回教示されたデータを溜め込み，オフラインで学習を行う．このとき，教示された時刻 t の集合を $T_o = \{t_1, t_2, ..., t_N\}$ とする．N は教示データ数である．時刻ごとの制御値，計測値および単語分割された発話文による複数の教示データから，モデルパラメータを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に，教示対象場所で文章発話を複数回行う．学習の際は，形態素解析器によって単語分割された発話文を音声認識単語 $O_{t,b}$ として与える．また，教示中の自己位置推定結果を固定ラグ平滑化処理 [7] した自己位置を x_t の初期値として用いる．一般に，平滑化を行うとオンライン推定よりも精度のよい推定値が得られることが知られている．位置分布は初期値は全て，μ_k= (一定の範囲内に一様乱数),$\Sigma_k = \begin{bmatrix} \sigma_{initial} & 0 \\ 0 & \sigma_{initial} \end{bmatrix}$ とする．

以下に，ギブスサンプリングを行う際の各要素ごとの事後分布を示す．

(12) 式は，位置分布の index i_t に関する事後分布である．

$$p(i_t = k \mid x_t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, C_t) \propto \mathcal{N}(x_t \mid \mu_{k=i_t}, \Sigma_{k=i_t})\mathrm{Mult}(i_t = k \mid \phi_{l=C_t}) \quad (12)$$

(13) 式は，場所概念の index C_t に関する事後分布である．このとき，$O_{t,\mathbf{B}}$ は時刻 t における発話文中の全ての単語を集めたものである．

$$\begin{aligned} &p(C_t = l \mid x_t, i_t, O_{t,\mathbf{B}}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, \pi) \\ &\quad \propto \mathrm{Mult}(O_{t,\mathbf{B}} \mid W_{l=C_t})\mathrm{Mult}(i_t = k \mid \phi_{l=C_t}) \\ &\qquad \times \mathrm{Mult}(C_t = l \mid \pi) \end{aligned} \tag{13}$$

場所の名前 $\mathbf{W}$ は，$l \in L$ ごとに (14) 式の様にサンプリングできる．このとき，β_{n_l} は事後パラメータであり，$\mathbf{O}_l$ は $t \in T_o$ の中で $C_t = l$ である発話文を集めたものである．

$$\mathrm{Dir}(W_l \mid \beta_{n_l}) \propto \mathrm{Mult}(\mathbf{O}_l \mid W_l)\mathrm{Dir}(W_l \mid \beta_0) \tag{14}$$

位置分布 $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ は，$k \in K$ ごとに (15) 式の様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，$\mathbf{x}_k$ は $t \in T_o$ の中で $i_t = k$ である教示位置を集めたものである．

$$\begin{aligned} &\mathcal{N}\text{-}\mathcal{W}(\mu_k, \Sigma_k \mid m_{n_k}, \kappa_{n_k}, V_{n_k}, \nu_{n_k}) \\ &\quad \propto \mathcal{N}(\mathbf{x}_k \mid \mu_k, \Sigma_k)\mathcal{N}\text{-}\mathcal{W}(\mu_k, \Sigma_k \mid m_0, \kappa_0, V_0, \nu_0) \end{aligned} \tag{15}$$

(16) 式は，場所概念の index の多項分布 π に関する事後分布である．

$$\mathrm{Dir}(\pi \mid C_{T_o}, \gamma) \propto \mathrm{Mult}(C_{T_o} \mid \pi)\mathrm{Dir}(\pi \mid \gamma) \tag{16}$$

位置分布の index の多項分布 ϕ_l は，$l \in L$ ごとに (17) 式の様にサンプリングできる．このとき，$\mathbf{i}_l$ は $t \in T_o$ の中で $C_t = l$ である位置分布の index を集めたものである．

$$\mathrm{Dir}(\phi_l \mid \mathbf{i}_l, \alpha) \propto \mathrm{Mult}(\mathbf{i}_l \mid \phi_l)\mathrm{Dir}(\phi_l \mid \alpha) \tag{17}$$

ロボットの自己位置のサンプリングに関しては，(18) 式，(19) 式の様に時刻 t に対する教示の有無でわける．

$$\begin{aligned} &p(x_t \mid x_{t-1}, x_{t+1}, u_t, u_{t+1}, z_t) \\ &\quad \propto p(x_{t+1} \mid x_t, u_{t+1})p(z_t \mid x_t)p(x_t \mid x_{t-1}, u_t) \\ &\qquad (t \notin T_o) \end{aligned} \tag{18}$$

$$\begin{aligned} &p(x_t \mid x_{t-1}, x_{t+1}, u_t, u_{t+1}, z_t, i_t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, C_t) \\ &\quad \propto p(x_{t+1} \mid x_t, u_{t+1})p(z_t \mid x_t)p(i_t \mid x_t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, C_t) \\ &\qquad \times p(x_t \mid x_{t-1}, u_t) \\ &\qquad (t \in T_o) \end{aligned} \tag{19}$$

3.5 場所概念学習後の自己位置推定

MCL の導出式の条件部に，t 時刻における発話認識文 $O_{t,\mathbf{B}}$ とモデルパラメータ集合 $\Theta = \{\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, \pi\}$ を加えた式を，(20) 式に示す．

$$\begin{aligned} &p(x_{0:t} \mid z_{1:t}, u_{1:t}, O_{1:t,\mathbf{B}}, \Theta) \\ &\quad \propto p(z_t \mid x_t)p(O_{t,\mathbf{B}} \mid x_t, \Theta)p(x_t \mid x_{t-1}, u_t) \\ &\qquad \times p(x_{0:t-1} \mid z_{1:t-1}, u_{1:t-1}, O_{1:t-1,\mathbf{B}}, \Theta) \end{aligned} \tag{20}$$

また，$p(O_{t,\mathbf{B}} \mid x_t, \Theta)$ に関しては，(21) 式の様に導出できる．

$$\begin{aligned} &p(O_{t,\mathbf{B}} \mid x_t, \Theta) \\ &\propto \sum_{C_t}\left[p(O_{t,\mathbf{B}}|W_{C_t}) \sum_{i_t}\left\{p(x_t|\mu_{i_t}, \Sigma_{i_t})p(i_t|\phi_{C_t})\right\}p(C_t|\pi)\right] \end{aligned} \tag{21}$$

このとき，$O_{t,\mathbf{B}}$ は，音声認識器の単語辞書に学習した言語モデルの全単語を加えた状態で，1-best 認識によって得る．

表 2: Phrase of each sentence

○○ だよ	○○ はこちらです
○○ です	こちらが ○○ になります
ここが ○○	このばしょが ○○ だよ
ここは ○○ です	このばしょのなまえは ○○
○○ にきました	ここのなまえは ○○ だよ

4. 実験

簡易な移動ロボットシミュレータを構築し，提案手法の有効性の検証を行う．音声認識器には大語彙連続音声認識システム Julius[*1] を利用した．Julius の単語辞書は，既存の大量語が登録された単語辞書を用いず，日本語音節のみを登録した単語辞書を使用する．マイクには，SHURE 社の PG27 USB を使用した．形態素解析器には，latticelm[*2] を使用した．

4.1 場所概念の学習

4.1.1 実験条件

座標原点は左上とし，x 軸は右方向，y 軸は下方向の 2 次元空間上で実験を行った．ロボットは前進，後進，右回転，左回転を行い 2 次元空間上を移動する．ロボット前方には複数の距離センサを持つ．距離センサはそれぞれ，センサ限界値以内に壁が存在する場合，壁までの距離を返す．センサ数は 20 個，センサ限界値は 150pixel とした．本実験での各パラメータ値は，$L = 10$，$K = 10$，$\alpha = 0.5$，$\gamma = 0.5$，$\beta_0 = 0.5$，$m_0 = [0, 0]^{\mathrm{T}}$，$\kappa_0 = 0.001$, $V_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\nu_0 = 2$，$\sigma_{initial} = 10000$ とし，イテレーション回数は，100 回とした．x_t についてはサンプリングを行わず，平滑化によって精度のよい推定値が得られているものと考え，近似としてロボットの真の座標を教示位置とする．学習対象の発話場所は，小さな四つの青い長方形の前付近とし，それぞれに対し 10 個の言い回しを含む合計 40 回分の発話教示を行った．教示する場所の名前はそれぞれ，“かいだんまえ” が 2 カ所と，“そうはつけん”，“ぷりんたあべや” である．各発話文における言い回しを表 2 に示す．

4.1.2 実験結果

学習結果の 1 例を以下に示す．位置分布を図示したものを図 3 に示す．黄色の各点群は，学習した位置分布に従う点を各位置分布に対して 500 個ずつ描画したものである．それぞれのふきだしは位置分布ごとの index 番号を示している．各場所概念における場所の名前を図 4 - 6 に，位置分布の index の多項分布を図 7 - 9 に示す．

この結果から，W_0 では “かいだんまえ” が最も確率が高く，ϕ_0 を見ると 0 番目と 2 番目の位置分布に対応していることがわかる．W_2 では “ぷりんぱべや” が最も確率が高く，ϕ_2 を見ると 3 番目の位置分布に対応していることがわかる．W_4 では “そはつけん” が最も確率が高く，ϕ_4 を見ると 1 番目の位置分布に対応していることがわかる．

5. おわりに

本稿では，以前の提案モデルを新たに拡張した場所概念獲得モデルの提案について述べた．

*1 使用バージョン : dictation-kit-v4.3.1-win GMM 版，
http://julius.sourceforge.jp/index.php
*2 使用バージョン : latticelm 0.4，
http://www.phontron.com/latticelm/index-ja.html

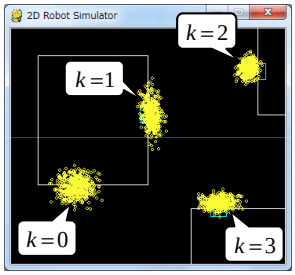

図 3: Learning result of the position distribution

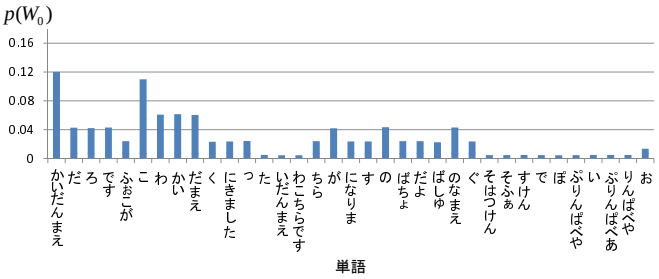

図 4: Name of location W_0

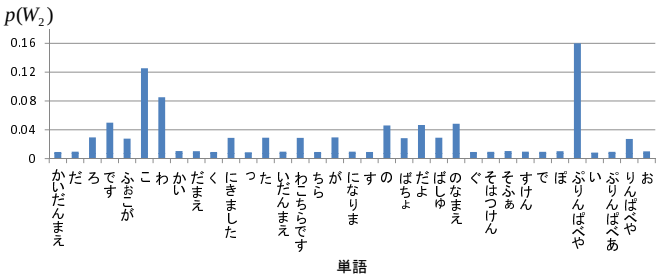

図 5: Name of location W_2

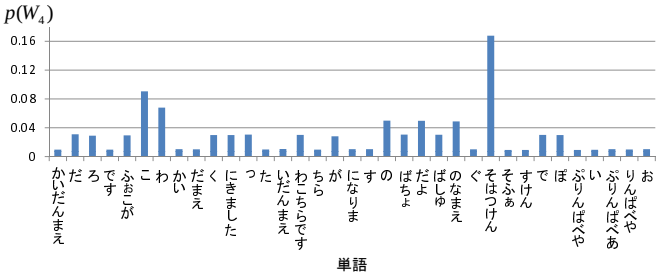

図 6: Name of location W_4

latticelm による教師なし形態素解析については，発話文全体に対して単語認識のゆらぎを抑える効果が見られたが，学習対象の場所の名前に対して細かく単語分割される場合があった．位置分布については，二つの学習対象場所を一つの位置分布が包含して学習される場合や，同じ学習対象場所に対して複数の位置分布に別れて学習される場合が見られた．場所の名前については，発話文全体に存在するような単語に対して場所概念が形成される場合が見られた．

また本研究では，環境の地図を与えた状態での自己位置推定を行ったが，SLAM(Simultaneous Localization And Mapping)[4] により事前に地図生成を行った後で本手法を適用することは可能であると考える．

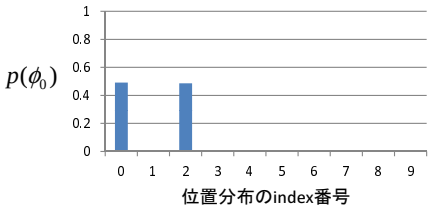

図 7: Multinomial distribution of index of the position distribution corresponding to W_0

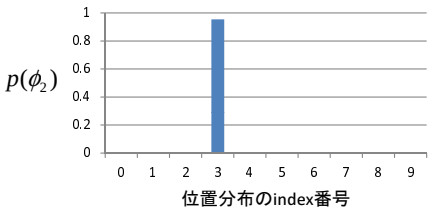

図 8: Multinomial distribution of index of the position distribution corresponding to W_2

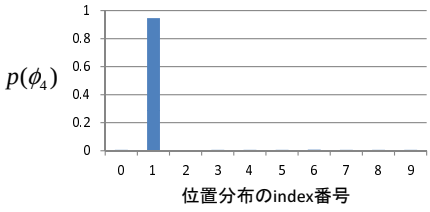

図 9: Multinomial distribution of index of the position distribution corresponding to W_4

参考文献

[1] 谷口彰, 吉崎陽紀, 稲邑哲也, 谷口忠大. 自己位置と場所概念の同時推定に関する研究. システム制御情報学会論文誌, Vol. 27, pp. 166–177, 2014.

[2] 田口亮, 岩橋直人, 船越孝太郎, 中野幹生, 能勢隆, 新田恒雄. 統計的モデル選択に基づいた連続音声からの語彙学習. 人工知能学会論文誌, Vol. 25, No. 4, pp. 549–559, 2010.

[3] 山田雄治, 服部公央亮, 田口亮, 梅崎太造, 保黒政大, 岩橋直人, 船越孝太郎, 中野幹生. 連続音声から場所の名前を学習する自律移動ロボット. 一般社団法人情報処理学会 全国大会講演論文集, Vol. 2011, No. 1, pp. 237–239, 2011.

[4] S. Thrun, W. Burgard, D. Fox, 上田隆一（訳）. 確率ロボティクス. 毎日コミュニケーションズ, 2007.

[5] Jayaram Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, Vol. 4, pp. 639–650, 1994.

[6] Graham Neubig, Masato Mimura, and Tatsuya Kawahara. Bayesian learning of a language model from continuous speech. *IEICE TRANSACTIONS on Information and Systems*, Vol. 95, No. 2, pp. 614–625, 2012.

[7] 北川源四郎. モンテカルロ・フィルタおよび平滑化について (特集 計算統計学の発展). 統計数理, Vol. 44, No. 1, pp. 31–48, 1996.