

自己位置と場所概念の同時推定に関する研究*

谷口 彰[†]・吉崎 陽紀[†]・稻邑 哲也[‡]・谷口 忠大[†]

Research on Simultaneous Estimation of Self-Location and Location Concepts*

Akira TANIGUCHI[†], Haruki YOSHIZAKI[†], Tetsunari INAMURA[‡] and Tadahiro TANIGUCHI[†]

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 integrate ambiguous speech recognition results with the model for self-localization on Bayesian approach. Experimental results show that a robot can obtain words for several locations and make use of them in self-localization task. In addition, we evaluate the performance of lexical acquisition task about words for places and show its effectiveness.

1. はじめに

サービスロボットなどの人間の生活環境で動作するロボットは、様々な環境において周囲の環境を認知し、人間とのインタラクションを通して、環境に対して人間が割り当てた語彙と、その語が指示する空間の領域を学んでいくことが重要である。この際、センサのノイズ、移動の誤差、音声認識誤りなど多くの不確実性への対処が重要である。

本研究では、事前に単語の知識を持たず、自己位置推定を行いながら環境を移動するロボットに、人が場所の名前を音声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。この目的には自己位置推定と語彙獲得が必要となるが、これらはそれぞれ以下のような問題を持つ。ロボットは自己位置推定を行う際、センサ情報から確率的な自己位置推定を行う。自己位置推定には推定誤差がつきものであり不確実である。複数の離れた位置にパーティクルが残る大域的な自己位置推定誤りの問題は、特に距離センサを用いた自己位置推定では頻繁に生じる。一方で、語彙獲得の問題では、単語

の知識を持たない状態で音声認識の結果が不確実な誤りありの音声から単語の同一性を推定するのは困難であるといえる。たとえば、音声認識結果“りんご”，“にんご”，“りんご”が同じかどうかを事前知識なしで判定するのは困難である。本論文では、ロボットの自己位置推定情報に不確実な音声認識結果を統合することで、これら二つの問題を同時に解決することを目指す。より具体的には、不確実な音声認識結果と自己位置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する。

2. 先行研究

2.1 語彙獲得

まず始めに、言語獲得や語彙獲得に関する先行研究について述べる。

中村らは、LDA(Latent Dirichlet Allocation)を拡張し聴覚、視覚、触覚などのマルチモーダル情報を統合し、物体概念を形成するマルチモーダルLDAを提案している[1]。さらに、マルチモーダルLDAによって形成された物体概念に単語を接地する手法を提案し、マルチモーダル情報と言語情報の統合を行っている[2]。これにより、物体に対する語彙獲得が行われ、物体から単語の推定、単語から物体の推定が可能となっている。

岩橋らは、ロボットが学習により人の発話を状況に応じて適切に理解し、また自然な発話を行わせることを目的とし、物体動作と文章の関係を獲得させる研究を行っている[3]。さらに、確率的な枠組みの中で、ロボット

* 原稿受付 2013年7月1日

† 立命館大学大学院 情報理工学研究科 Graduate School of Information Science and Engineering, Ritsumeikan University; 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, JAPAN

‡ 国立情報学研究所 National Institute of Informatics; 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN
Key Words: self-localization, ambiguous speech recognition, lexical acquisition.

に人との発話や行動のインタラクションを通して言語コミュニケーション能力をオンラインで学習させるシステムを構築している[4,5]。

文献[4]の発展として、杉浦らの物体操作対話をタスクとした研究も行われている[6]。この研究では、言語によるオブジェクト参照、動作参照のあいまい性の解消を行い、言語獲得ロボットに発話理解確率の推定により動作や発話を生成させる手法を提案している。

J. Hornstein らの研究では、子供の言語獲得プロセスを模倣したヒューマノイドロボットにおける言語獲得のための手法を提案している[7]。聴覚、視覚といったマルチモーダル情報を利用し、人との相互作用によって言語獲得を行う。あらかじめプログラムされた言語知識を必要とせず、単語や音素などの言語構造をパターン認識や階層的なクラスタリングにより獲得している。

D. Roy らの研究では、単一の物体の画像および人間と幼児の対話音声から、音声のセグメンテーション、単語の探索、視覚的な分類を行うモデルを提案し、物体の名前の学習を行っている[8]。また、状況に依存した自然な発話音声における、音響や文法に関するあいまいさと文脈や意図に関するあいまいさの両方の側面を考慮したフレームワークが提案されている[9]。実験では、日常的なオブジェクトのあるゲーム内環境での操作や会話から得られたデータが用いられ、結果として指示対象物の決定における改善が示されている。

S. Qu & J. Y. Chai の研究では、音声と視線の共起に注目し、対話システムにおける語彙の問題に取り組むため、自動的に新しい単語を獲得する教師なし学習のアプローチを提案している[10,11]。この提案手法ではベースモデルとして IBM 翻訳モデル[12]を利用し単語とエンティティの関連付け確率の推定を行っている。さらに単語獲得におけるユーザの言語行動、ドメイン知識、対話コンテキストを組み込むことでベースモデルの改良を行っている。文献[11]では、仮想世界での混合主導権型の状況依存対話により得られたデータを元に、ベースモデルおよび各改良手法において単語獲得性能の比較評価を行っている。

以上のように、物体を対象とした言語獲得、語彙獲得の研究は多く行われている。しかし、誤りあり音声認識を利用した場所に関する語彙獲得の研究はあまり行われてこなかった。

2.2 場所と語彙の同時学習

以下では、本研究により関連した場所と語彙に関する先行研究について述べる。

A. S. Huang らの、事前にマッピングしラベル付けをした環境で、マイクロエアビークルを自然言語命令により動作させることを目的とした研究がある[13]。この研究では、方向に関する事柄に注目し、自然言語による多様な命令を理解するシステムが構築されているが、地図

上のラベル自体の学習は行われていない。

本田らの研究では、ラットの海馬に存在する場所細胞を半教師あり学習可能なニューラルネットワークによって実装し、ロボットに自律移動を行わせる手法が提案されている[14]。入力は、距離センサから取得される環境の形状を表す特微量と環境の形状を表すラベルであり、ラベルは言語インタラクションに利用される。提案手法により、言語インタラクションにおいて動作指示、目的地指示、現在位置説明、経路説明が可能となると述べられている。しかし、テキスト入力により、あらかじめ決められたラベル名が与えられているため、人の発話音声による場所の名前の学習は行われていない。音節認識誤りを含む認識結果からの未知語の獲得と、誤りのない入力からの学習では理論的に大きな隔たりがある。

これに対して、単語知識のないロボットに、多様な言い回し発話から単語の正しい分節、音素系列、および、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[15]。山田らの研究では、連続ベクトルを指示対象として扱えるように先の手法[15]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使う手法が提案されている[16]。任意の場所での発話教示による学習後、学習に用いていない位置情報を入力し、場所に対応した正しい単語の出力をしている。しかし、獲得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利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入力を前提にしたうえで、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、場所をある空間的な広がりを持ったものとし、位置を環境中のある特定の座標や局所的な地点のこととして、語を使い分ける。本研究では、大域的な不確かさにロボットが対応しなければならない場合などによく用いられ、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 Localization) に場所概念を導入し拡張したモデルを提案する。場所概念には、場所の名前とその名前によって表現される場所の分布という二つの知識が含まれるものとする。ロボットは事前に単語の知識を持たず、日本語の音節のみを認識できるものとする。

3.1 提案モデルとタスクの概要

地図のある環境上で移動ロボットに建物内を移動させ、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 Fig. 1 を例に説明する。Fig. 1 (a) のように、三つの物体前の場所を学習対象の場所とする。たとえば、Fig. 1 (b) のように、人とロボットがテレビの前にいる場合、人がロボットにその場所の名前“てれびまえ”と発話し教示を行う。白い棚付近に移動した場

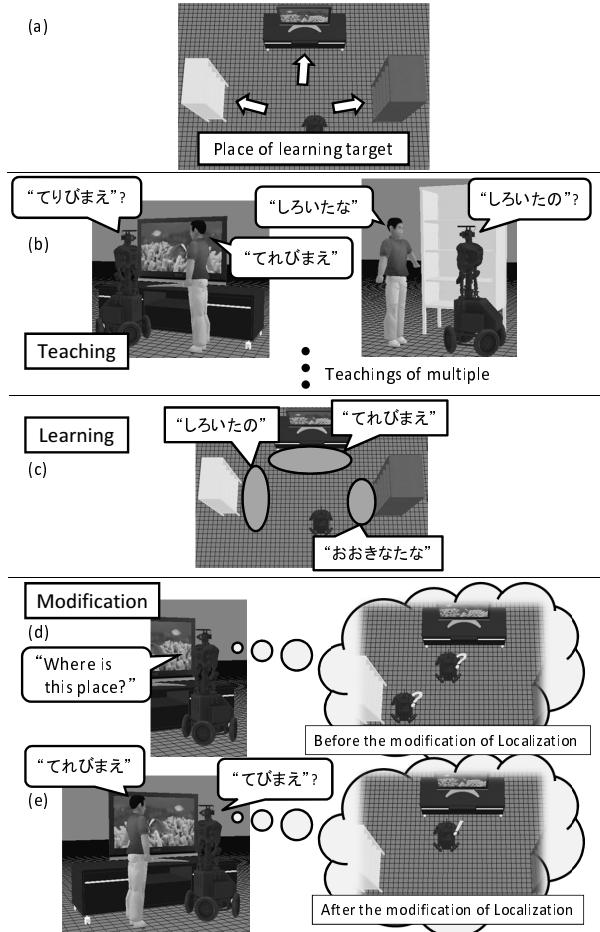


Fig. 1 Schematic diagram of the proposed method

合、人がロボットに“しろいたな”と教示する。大きな棚の前に来たときも同様である。このような教示を複数回行った後、ロボットは場所概念の学習を行う。Fig. 1 (c) のように、教示した各物体前付近に場所の分布が構成され、その分布に対応した場所の名前が学習される。この場合、ロボットは誤りありの音声認識を行うため、“しろいたの”のような、誤りの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ながら移動している。Fig. 1 (d) のように、ロボットは実際にはテレビの前にいるのだが、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。Fig. 1 (e) のように、人は“てれびまえ”であると発話する。するとロボットは、“てびまえ”と誤認識するものの、発話された場所の名前と学習した場所概念を利用して、自分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。

本研究ではおもに、(1) 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と、(2) 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情報の修正について問題とする。

本提案手法による自己位置情報の修正が有効な場合

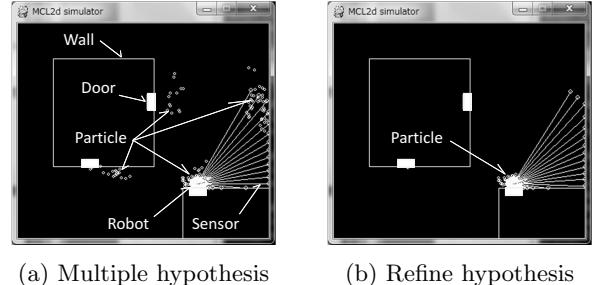


Fig. 2 Valid example of the modification of Localization

を考える。ロボットが自己位置推定を行っているとき、Fig. 2 (a) のように複数の離れた場所にパーティクルが分布している状況が起こりうる。この原因は、主に空間の回転対称性や局所的に同一な特徴の存在する環境などにより、異なった位置で似通ったセンサ値が取得されることによるものである。提案手法では、環境中に局所的に存在する場所概念を利用することでこの問題の解決を図っている。Fig. 2 (a) の小さな三つの白い長方形はドアを表しており、このドア前付近に場所概念が学習されているものとする。このとき、ロボットに場所の名前を発話すると、Fig. 2 (b) のようにパーティクルがロボットの真の位置に近い分布になり自己位置情報の修正が行われる。

3.2 場所概念の定義

学習により獲得された場所概念の集合 λ を(1)式で表す。 L は場所概念の数である。 i 番目の場所概念は、(2)式のように定義する。

$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_L\} \quad (1)$$

$$\lambda_i = \{W_i, \mu_i, \Sigma_i\} \quad (2)$$

このとき、 $W_i = (l_1^i l_2^i \dots l_{d_i}^i)$ は学習により得た i 番目の場所の名前である。 l_j^i は W_i の j 番目の文字であり、 d_i は W_i の文字数である。場所の名前の集合 \mathbf{W} は、(3)式のように表せる。また、場所の分布は多変量ガウス分布を仮定する。そのため、 μ_i は i 番目の分布の平均、 Σ_i は i 番目の分布の共分散行列である。平均の集合 $\boldsymbol{\mu}$ は(4)式、共分散行列の集合 $\boldsymbol{\Sigma}$ は(5)式のように表せる。

$$\mathbf{W} = \{W_1, W_2, \dots, W_L\} \quad (3)$$

$$\boldsymbol{\mu} = \{\mu_1, \mu_2, \dots, \mu_L\} \quad (4)$$

$$\boldsymbol{\Sigma} = \{\Sigma_1, \Sigma_2, \dots, \Sigma_L\} \quad (5)$$

時刻 t における、ロボットが聞き取った音声情報 O_t に対応した場所概念の認識状態を C_t で表す。 C_t は、場所概念の認識状態の集合を \mathbf{C} としたとき、(6)式のように 1 から L までの離散値をとる。

$$C_t \in \mathbf{C} = \{1, 2, \dots, L\} \quad (6)$$

つまり、 C_t 番目の場所概念は、 $\lambda_{C_t} = \{W_{C_t}, \mu_{C_t}, \Sigma_{C_t}\}$ と表せ、 $W_{C_t}, \mu_{C_t}, \Sigma_{C_t}$ は、それぞれ C_t 番目の場所概念

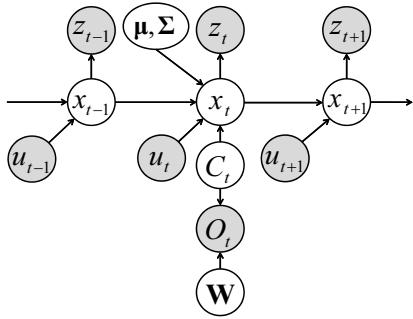


Fig. 3 Graphical model of the proposed method

Table 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	場所概念の認識状態
O_t	音声認識単語
\mathbf{W}	場所の名前の集合
μ, Σ	場所の分布（平均、共分散行列）の集合

の場所の名前、場所の分布の平均、共分散行列である。

3.3 場所概念を導入した自己位置推定

MCL に場所概念の認識状態 C_t 、場所の分布 (μ, Σ) と場所の名前 \mathbf{W} および音声認識単語 O_t の関係を加えたグラフィカルモデルを Fig. 3 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 Table 1 に示す。

MCL の導出式に O_t を加えた式を、(7) 式に示す。

$$\begin{aligned} p(x_{0:t} | z_{1:t}, u_{1:t}, O_{1:t}) \\ \propto p(z_t | x_t) p(O_t | x_t) p(x_t | x_{t-1}, u_t) \\ \times p(x_{0:t-1} | z_{1:t-1}, u_{1:t-1}, O_{1:t-1}) \end{aligned} \quad (7)$$

(7) 式で導出された $p(O_t | x_t)$ について、グラフィカルモデルの関係により導出したものを、(8) 式に示す。(8) 式は、 x_t という位置で O_t という単語を聞き取ったときの確率を表しており、計測モデル $p(z_t | x_t)$ と同じく尤度として計算を行う。

$$\begin{aligned} p(O_t | x_t) &= \sum_{C_t} p(O_t | C_t) p(C_t | x_t) \\ &\propto \sum_{C_t} p(O_t | \mathbf{W}, C_t) p(x_t | \mu, \Sigma, C_t) \\ &\quad \times p(C_t) \end{aligned} \quad (8)$$

また本研究では、 $p(C_t)$ は無情報を仮定し、 $p(C_t) = 1/L$ のように一様分布とする。

$$p(O_t | \mathbf{W}, C_t) \propto \exp(-\beta \text{LD}(O_t, W_{C_t})) \quad (9)$$

$p(O_t | \mathbf{W}, C_t)$ については、編集距離によって場所の名前と認識文字の近さを表す確率の式として、(9) 式のよ

うに定義する。 β は編集距離の値による影響度合いを確率式に反映するパラメータであり、LD は編集距離を計算する関数である。編集距離とは、二つの文字列の異なり具合を量量化するものであり、文字の挿入・削除・置換により一方を他方に変形するまでの最小手順回数を数えたものである。

$$\begin{aligned} p(x_t | \mu, \Sigma, C_t) \\ = \frac{1}{(2\pi)^D/2 |\Sigma_{C_t}|^{1/2}} \\ \times \exp\left(-\frac{1}{2}(x_t - \mu_{C_t})^T \Sigma_{C_t}^{-1} (x_t - \mu_{C_t})\right) \end{aligned} \quad (10)$$

$p(x_t | \mu, \Sigma, C_t)$ については、場所の分布と自己位置情報との近さを表す確率の式として、(10) 式のように定義する。この式は、次元数 D の多変量ガウス分布を表す。

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 $T_o = \{t_1, t_2, \dots, t_N\}$ とする。 N は教示データ数である。発話位置と認識単語による複数の教示データから場所の名前と場所の分布をギブスサンプリングによって推定する。場所概念の数 L は既知とする。ギブスサンプリングによる場所概念学習の処理のイメージ図を Fig. 4 に示す。各ウインドウ枠は場所概念の学習の遷移を各ステップごとに表したものであり、ウインドウ内部にある大きな四角形の枠は壁を表している。この図の場合、ステップ 1 のウインドウ内部の小さな三つの白い長方形で表した部分を学習対象の場所とする。場所概念学習アルゴリズムを Algorithm 1 に示す。各ステップの詳細については付録 1. に記載した。

4. 実験 I：簡易シミュレータによる実験

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。場所概念の学習および学習した場所概念について自己位置推定と語彙獲得に関しての評価実験を行う。

本研究では、音声認識器に大語彙連続音声認識システム Julius[17]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。これによって、発話音声を音節ごとに分け音節列として認識することができる。マイクには、SHURE 社の PG27 USB 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 Fig. 5 上で実験を行った。Fig. 5 の小さな三つの白い長方形はドアを表しており、このドア前付近を学習対象の場所とする。ロボットは前進、後進、右回転、

¹ インストールバージョン：dictation-kit-v4.2-win.

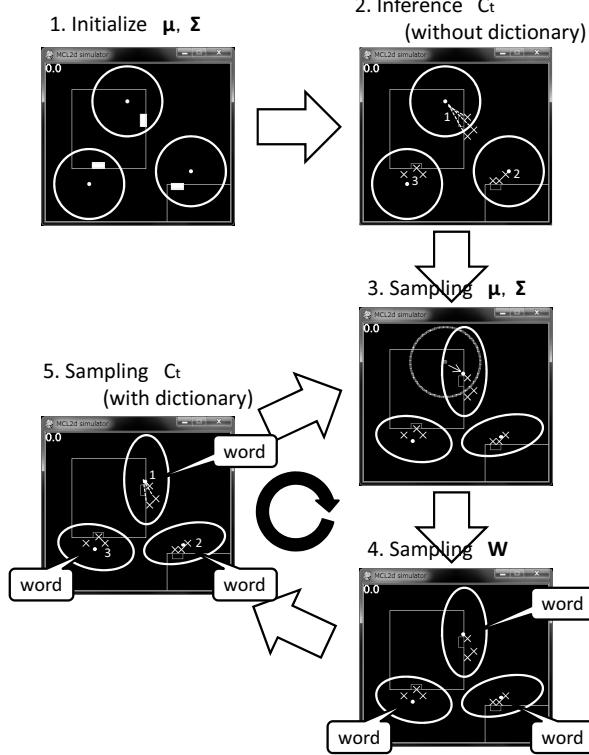


Fig. 4 Flow diagram of learning of Location Concepts by Gibbs sampling

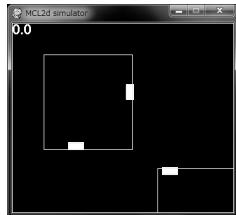


Fig. 5 Environment to be used for learning

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。本実験では、場所概念の数は $L=3$, $\sigma_{initial}=10000$ 、ステップ(3)における事前分布のハイパーパラメータの初期値は、 $\beta_0=0.001$, $m_0=[0,0]^T$, $V_0=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\nu_0=1$ とし、ステップ(6)での繰り返し回数は、10回とした。

本論文では、ロボットの自己位置推定の情報は用いず、近似としてロボットの真の座標を教示データの位置情報とする。この設定が近似として妥当であると考える理由を以下に述べる。自己位置推定のタスクは通常、オンラインであるが、場所概念の学習はオフラインで行う。このため、発話教示された時点(t_n)の自己位置は、時刻 t までの観測データ($z_{1:t}$)に基づくオンライン自己位置推定ではなく、タスクの完了後、すべての観測データ($z_{1:T}$)

Algorithm 1 The Learning by Gibbs Sampling

```

 $\mathbf{X} = \{x_{t_1}, x_{t_2}, \dots, x_{t_N}\}$ 
 $\mathbf{O} = \{O_{t_1}, O_{t_2}, \dots, O_{t_N}\}$ 
procedure Location_Concept_Learning( $\mathbf{X}, \mathbf{O}, L, N$ )
     $C_{\mathbf{T}_o} = \{C_{t_1}, C_{t_2}, \dots, C_{t_N}\}$ 
    iteration_number : イテレーション回数
    // (1)  $\mu, \Sigma$  を初期化
     $\mu, \Sigma = \text{Initialize}()$ 
    // (2)  $\mathbf{W}$  の情報無しで、 $C_t$  をサンプリング
    for  $t=1$  to  $N$  do
         $C_t \sim p(x_t | \mu_{C_t}, \Sigma_{C_t})$ 
    end for
    // (6) (3)～(5)の過程を複数回繰り返す
    for  $i=1$  to iteration_number do
        // (3)  $\mu, \Sigma$  をサンプリング
        for  $c=1$  to  $L$  do
             $\mu_c, \Sigma_c \sim \left( \prod_{\substack{c=C_t \\ t \in \mathbf{T}_o}} p(x_t | \mu_{C_t}, \Sigma_{C_t}) \right) p(\mu_c, \Sigma_c)$ 
        end for
        // (4)  $\mathbf{W}$  をサンプリング
        for  $c=1$  to  $L$  do
             $W_c \sim \left( \prod_{\substack{c=C_t \\ t \in \mathbf{T}_o}} p(O_t | W_{C_t}) \right) p(W_c)$ 
        end for
        // (5)  $\mathbf{W}$  の情報有りで、 $C_t$  をサンプリング
        for  $t=1$  to  $N$  do
             $C_t \sim p(x_t | \mu_{C_t}, \Sigma_{C_t}) p(O_t | W_{C_t})$ 
        end for
    end for
    return  $\mathbf{W}, \mu, \Sigma$ 
end procedure

```

発話単語	認識単語				
そうはつけん	ぶはつけん	そふつきゅ	そつけん	すがつけん	そぶはつけん
しょむはつけん	しょむはつけん	そつけん	そつけん	ぶはつけん	そぼはつけん
かいだん	かりだん	かいだん	かいだん	かにたん	かにだん
かいだん	かいだん	かいだん	かにぶん	かいるん	かにだん
ちのうけん	ちのけん	しのけん	きのおけん	ちのうけん	ちのけん
ちのうけん	ちのけん	ちのけん	ちのけん	ちのけん	ちのけん

Fig. 6 Julius recognition results using the Japanese syllable word dictionary (Three words)

が与えられた状態から推定することができる¹。このような観測データに基づく推定を行えば、十分に精度の良い自己位置情報を得ることができると考えられる。

発話位置は、それぞれ三つのドア前付近とし、10回ずつ、合計30回分の教示データを与えた。発話単語はそれぞれ、“そうはつけん”, “かいだん”, “ちのうけん”とし、教示データの言語情報には、Fig. 6の各発話単語に対する認識結果を利用した。

¹これは平滑化アルゴリズムによって実現できる[18]。一般に、平滑化を行うとオンライン推定よりも比較的精度のよい推定値が得られる。

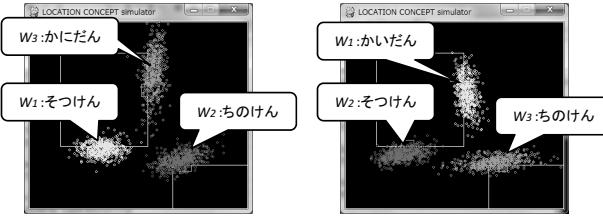


Fig. 7 Learning result 1

Fig. 8 Learning result 2

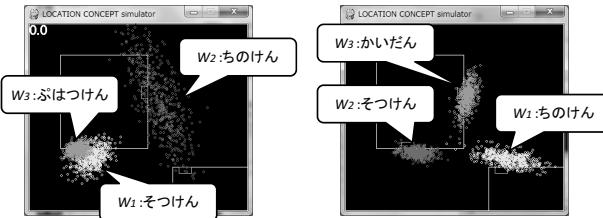


Fig. 9 Learning result 3

Fig. 10 Learning result 4

4.1.2 実験結果

15回分の試行による学習結果から、4例を図示したものをFig. 7 - 10に示す。各点群は、学習した場所の分布に従う点を各場所概念に対して500個ずつ描画したものである。 $W_1 \sim W_3$ は、それぞれの分布に対応した場所の名前である。今回は、三つのドア前に対応した三つの場所概念を学習させることを行った。15回分の学習結果中、2回のみでFig. 9のように明らかに、各ドアに対して正しく1対1対応で学習できていない結果が見られた。場所概念の学習に用いている混合分布の推定では、局所解に陥ってしまう可能性がある。このため、Fig. 9のような結果が見られるのは、局所解に陥っていることが原因として考えられる。その他の学習結果では、Fig. 7, Fig. 8, Fig. 10のような学習結果が得られ、多くの場合で三つのドア前に場所概念が学習できていることがわかる。

4.2 学習した場所概念の評価のための比較実験

4.2.1 実験条件

つぎに、先の場所概念の学習結果を用いた自己位置情報の修正について確認した。学習時と同じ2次元空間上で通常のMCLと提案手法によって場所概念を学習した状態のMCLとで自己位置の推定誤差精度の比較を行った。パーティクル数は $M = 300$ とし、ロボットの移動範囲内(1~350, 1~300)に一様に分布とした。ロボットは1stepごとに1回の制御動作を行う。ロボットは300step移動し、各ドア前を“かいだん”, “ちのうけん”, “そはつけん”と順番に巡るような動作をする。このとき、すべての試行において、ロボットは同じ動作とした。提案手法でのみ、それぞれドア前付近に来たとき場所の名前を教示した。教示単語は、Fig. 6の各ドアに対する認識単語からランダムに選ぶ。

MCLにおける推定誤差の評価には、上田ら[19]の評価方法を参考にした。まず、stepごとに xy 平面での推

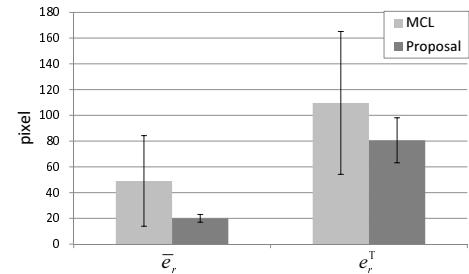


Fig. 11 Comparison of the estimated error due to MCL and the proposed method (Experiment I)

定誤差((11)式)を記録する。

$$e_t = \sqrt{(\bar{x}_t - x_t^*)^2 + (\bar{y}_t - y_t^*)^2} \quad (11)$$

このとき、 $\bar{x}_t = \sum_{i=1}^M w_t^{(i)} x_t^{(i)}$, $\bar{y}_t = \sum_{i=1}^M w_t^{(i)} y_t^{(i)}$ であり、 x_t^*, y_t^* はロボットの真の位置座標である。 $w_t^{(i)}, x_t^{(i)}, y_t^{(i)}$ はそれぞれ、時間 t における*i*番目のパーティクルの重み、 x 座標、 y 座標である。試行後、 e_t の平均値 \bar{e}_r と、区間 $[0, \gamma]$ が e_t の95%以上を含む γ の最小値 e_r^T を計算する。 e_r^T は推定の安定性を示す指標として見なせる。

4.2.2 実験結果

15回分の実験結果を \bar{e}_r , e_r^T によって比較したグラフを、Fig. 11に示す。どちらの指標でも、推定誤差は提案手法の方が小さいという結果が得られた。また、この結果で有意水準5%の片側t検定を行ったところ、 \bar{e}_r , e_r^T ともに有意差が見られた。この結果は、自己位置推定中、複数の離れた場所に仮説が存在した場合に、場所概念を導入することにより実際にロボットがいる位置付近の仮説のみに絞り込むことができたからだと考えられる。よって、学習により大域的な自己位置情報の修正を行えるような場所概念が獲得できていることが確認できた。

4.3 語彙獲得の評価

4.3.1 実験条件

場所概念の学習において、データごとに場所概念の認識状態 C_t が推定される。場所概念の認識状態の最終的な推定結果を見ることで、場所概念推定についての評価を行う。人間による分類と場所概念の学習による C_t の推定結果の正解率を見る。人間による分類は、どの教示データがどの場所概念として教示されたかを正解とし、Fig. 12のように表す。Fig. 12の縦軸は教示データを並べたものであり、縦軸に表された数字は教示データの番号を表す。横軸は教示データに対応した場所概念の認識状態を表す。このとき、1は“そはつけん”，2は“かいだん”，3は“ちのうけん”とする。図の白い部分は、該当するデータがその場所概念として教示(学習の場合は推定)されたことを意味する。 C_t 推定結果と人間による分類との一致率を、推定正解率(Estimation Accuracy Rate: EAR)として、(12)式に示す。

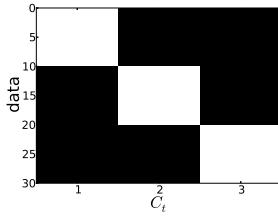


Fig. 12 Classification by human (Correct answer)

$$\text{EAR} = \frac{\text{推定正解データ数}}{\text{全表示データ数}} \times 100 \quad (12)$$

提案手法である、位置情報 x_t と言語情報 O_t を使用した場所概念の学習と、言語情報 O_t のみ (位置情報 x_t なし) とした場合での場所概念の学習において、場所概念の推定正解率を比較する。つまり、言語情報のみによる語彙獲得と、言語情報に位置情報を付加したときの語彙獲得の精度について評価を行うことを意味する。

言語情報 O_t のみの場合における、場所概念の学習ステップ (提案手法との差異) を以下に示す。

(1) μ, Σ を初期化

提案手法と同様。

(2) \mathbf{W} の情報なしで、 C_t をサンプリング

(13) 式のように $t \in T_o$ についてサンプリングする。

$$C_t \sim p(C_t) \quad (13)$$

(3) μ, Σ をサンプリング

(14) 式のように $c \in \mathbf{C}$ についてサンプリングする。

$$\mu_c, \Sigma_c \sim p(\mu_c, \Sigma_c) \quad (14)$$

(4) \mathbf{W} をサンプリング

提案手法と同様。

(5) \mathbf{W} の情報ありで、 C_t をサンプリング

(15) 式のように $t \in T_o$ についてサンプリングする。

$$C_t \sim p(O_t | W_{C_t}) p(C_t) \quad (15)$$

(6) (3)～(5) の過程を複数回繰り返す。

4.3.2 実験結果

位置情報 x_t と言語情報 O_t をともに使用した場合と、言語情報 O_t のみを使用した場合において、それぞれ 10 回分の学習を行わせた。 O_t のみを使用した場合の学習による推定結果の平均を図示したものを Fig. 13(a) に示す。また、 x_t と O_t を使用した場合の学習による推定結果の平均を図示したものを Fig. 13(b) に示す。Fig. 13(a), (b) の凡例 Fig. 14 のように、データごとの C_t 推定回数を輝度の大きさで表している。10 回分の学習による推定正解率の結果を Table 2 に示す。言語情報 O_t のみの場合よりも、位置情報 x_t と言語情報 O_t を使用した場合の方が高い推定精度であった。この結果より、語彙獲得において、言語情報に位置情報を付加することによって、より正しい語彙獲得が行われることが確認された。

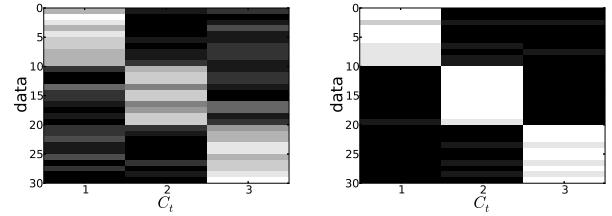
Fig. 13 Estimation results of C_t (Experiment I)

Fig. 14 Legend of the gray scale

Table 2 Result of EAR (Experiment I)

学習条件	平均 EAR (標準偏差)
Only O_t	76.67%(4.71)
Proposed	96.67%(9.56)

5. 実験 II : SIGVerse 上での実験

エージェントシミュレータ SIGVerse[20] を用い、より現実的な環境上で提案手法の有効性の検証を行う。SIGVerse とは、サーバクライアント形式の社会的知能発生学シミュレータである。SIGVerse は 3 次元空間上で、力学計算、知覚シミュレーション、コミュニケーションのシミュレーションを同時に実行することができる。

場所概念の学習および学習した場所概念について自己位置推定と語彙獲得に関する評価実験を行う。音声認識器およびマイクには、4. で使用したものと同様のものを用いる。

5.1 場所概念の学習

5.1.1 実験条件

3 次元空間上に Fig. 15 で示すような環境を作成し実験を行った。Fig. 15 の小さな四つの白い三角形はランドマークの位置を表している。本実験では、場所概念の数は $L = 4$ とし、 $\sigma_{initial} = 10000$ 、ステップ (3) における事前分布のハイパーパラメータの初期値は、 $\beta_0 = 0.001$ 、

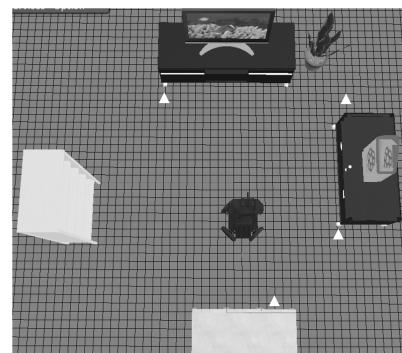


Fig. 15 Environment to be used for learning and landmark position

発話単語	認識単語				
	しろいたな	しにたな	しろいたが	しのたが	しめたな
ちにたの	ちにたの	つたな	しにたな	しのたな	しるいたな
つくえ	つくべ	つくね	つくう	つくえ	つくめ
	つくね	つくけ	つつね	つくね	つくめ
ごみばこ	ごみわこ	ぐみわこ	ぐみわこ	ぐみわこ	くみわこ
	ごみわこ	ぐみわこ	ぐみわこ	ぐみばこ	ぐみばく
てれびまえ	てりまえ	てびまえ	てぎくなえ	てれなえ	てりまえ
	てれなえ	てづりまえ	てりまえ	ててぎなえ	てりびなえ

Fig. 16 Julius recognition results using the Japanese syllable word dictionary (Four words)

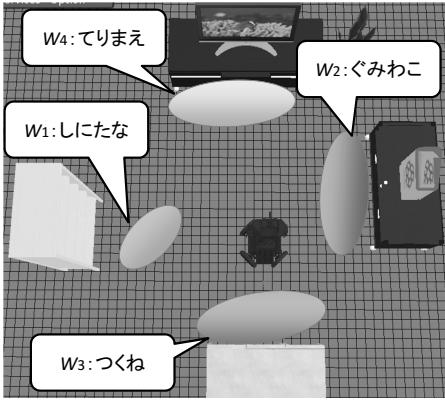


Fig. 17 Learning result on SIGVerse

$m_0 = [0, 0]^T$, $V_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\nu_0 = 1$, ステップ(6)での繰り返し回数は、10回とした。発話位置は、四つの物体前付近とし、それぞれ10回ずつ合計40回分の教示データを与えた。発話単語はそれぞれ、“しろいたな”, “つくえ”, “ごみばこ”, “てれびまえ”とし、教示データの言語情報には、Fig. 16 の各発話単語に対する認識結果を利用した。

5.1.2 実験結果

学習結果から、1例を図示したものを Fig. 17 に示す。各楕円は、学習した場所の分布を表す。 $W_1 \sim W_4$ は、それぞれの分布に対応した場所の名前である。この結果からも、四つの物体前の場所に対応した四つの場所概念が学習できていることがわかる。

5.2 学習した場所概念の評価のための比較実験

5.2.1 実験条件

つぎに、先の場所概念の学習結果を用いた自己位置情報の修正について確認した。SIGVerse 上でランドマーク型の MCL を実装し、通常の MCL と場所概念を学習した状態の MCL とで自己位置の推定誤差精度の比較を行った。ロボットのカメラ視野角内にランドマークがある場合、ランドマークを認識する。センサ情報として、認識したランドマークまでの距離と角度を得る。カメラはロボット前方に設置しており、カメラ視野角は45度である。パーティクル数は $M = 1000$ とし、初期パーティクルは一様に分布した。ロボットは1 step ごとに1回の

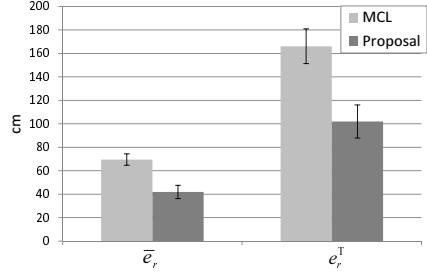


Fig. 18 Comparison of the estimated error due to MCL and the proposed method (Experiment II)

制御動作を行う。ロボットは400 step 移動し、各物体前の場所を“しろいたな”, “つくえ”, “ごみばこ”, “てれびまえ”と順番に巡るような動作をする。このとき、すべての試行において、ロボットは同じ動作とした。提案手法でのみ、各物体前付近の場所に来たとき場所の名前を教示した。教示単語は、Fig. 16 の各発話単語に対する認識結果からランダムに選ぶ。推定誤差の評価方法は、4.2.1 と同様である。

5.2.2 実験結果

10回分の実験結果を \bar{e}_r , e_r^T によって比較したグラフを、Fig. 18 に示す。ランドマーク型の MCL を用いた場合でも距離センサを用いた 4.2.2 と同様、推定誤差は提案手法の方が小さく、有意水準 5% の片側 t 検定により、 \bar{e}_r , e_r^T ともに有意差が見られた。よって、自己位置情報の修正を行えるような場所概念が獲得できていることが確認でき、また距離センサを用いた場合だけでなくランドマーク型の MCL でも提案手法は有効であることが示せた。

5.3 語彙獲得の評価

5.3.1 実験条件

4.3 と同様の実験を行った。人間による分類は、Fig. 12 と同様の形で表し、1は“しろいたな”，2は“つくえ”，3は“ごみばこ”，4は“てれびまえ”とする。

5.3.2 実験結果

位置情報 x_t と言語情報 O_t をともに使用した場合と、言語情報 O_t のみを使用した場合において、それぞれ 10 回分の学習を行わせた。 O_t のみを使用した場合の学習による推定結果の平均を図示したものを Fig. 19(a) に示す。また、 x_t と O_t を使用した場合の学習による推定結果の平均を図示したものを Fig. 19(b) に示す。10回分の学習による推定正解率の結果を Table 3 に示す。

言語情報 O_t のみの場合よりも、位置情報 x_t と言語情報 O_t を使用した場合の方が高い推定精度であった。この結果においても、言語情報に位置情報を付加することによって、より正しい語彙獲得が行われることが確認された。

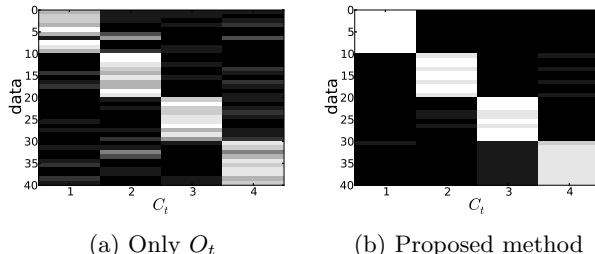
Fig. 19 Estimation results of C_t (Experiment II)

Table 3 Result of EAR (Experiment II)

学習条件	平均 EAR (標準偏差)
Only O_t	80.75%(10.98)
Proposed	95.50%(6.57)

6. 実験 III：異なった場所での同単語発話による場所概念の学習

ある空間において、異なった場所で同じである場所の名前が存在する場合でも、本手法では、教示される場所が大きく異なれば異なる場所概念として獲得される。これを調べるために、簡易シミュレータを使用し、場所概念の学習と語彙獲得の評価を行う。同じ場所の名前を教示したとしても異なった場所に場所概念が獲得できるかどうかを調べる。

音声認識器およびマイクには、4で使用したものと同様のものを用いる。

6.1 場所概念の学習

6.1.1 実験条件

場所概念の学習の条件は、一つのドアの位置と発話単語以外すべて4.1.1と同じとした。左上側のドアと右下側のドアに対して、ともに同じ場所の名前を教示する。発話単語は、“そうはつけん”と“じっけんしつ”とし、“じっけんしつ”に関しては左上側のドアと右下側のドアに対して教示する。“そうはつけん”については、Fig. 6の認識結果を、“じっけんしつ”については、Fig. 20の認識結果を利用した。

6.1.2 実験結果

学習結果の1例をFig. 21に示す。この結果から、場所の名前が同じであっても、異なった場所に場所概念が獲得できていることがわかる。

6.2 語彙獲得の評価

6.2.1 実験条件

4.3と同様の実験を行った。人間による分類は、Fig. 12と同一である。1は“そうはつけん”，2は“じっけんしつ”(左上側)，3は“じっけんしつ”(右下側)とする。

6.2.2 実験結果

それぞれ10回分の学習を行わせた。 O_t のみを使用した場合の学習による推定結果の平均を図示したものをFig. 22(a)に示す。また、 x_t と O_t を使用した場合の学

発話単語	認識単語				
	じっけんしつ (左上側)	じっくんしつ じきょちつ	じけんしつ じけんちつ	ちっくんしつ じっくんしつ	ちっけんしつ じっけんちつ
じっけんしつ (右下側)	じっけんしつ じっくんしつ	じとんちつ じつきゅんしつ	じきんつつ ちつきゅんしつ	ちけんつ じけん	ちっくんしつ ちっけんす

Fig. 20 Julius recognition result of words spoken

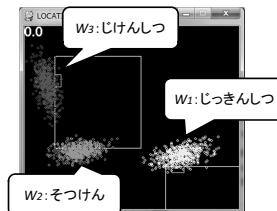


Fig. 21 Learning result

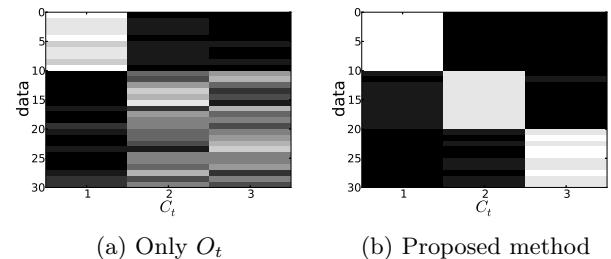
Fig. 22 Estimation results of C_t (Experiment III)

Table 4 Result of EAR (Experiment III)

学習条件	平均 EAR (標準偏差)
Only O_t	66.00%(7.67)
Proposed	94.67%(14.67)

習による推定結果の平均を図示したものをFig. 22(b)に示す。10回分の学習による推定正解率の結果をTable 4に示す。

結果として、Fig. 22(a)では、2番目と3番目の場所概念において、“じっけんしつ”と発話したデータを混同する場合が見られるのに対し、Fig. 22(b)では、より人間による分類に近い推定が行えていることがわかる。よって、場所の名前が同じ場合でも位置情報を付加することによって、より場所を識別するという意味で正確な語彙獲得が行われることが確認された。

7. おわりに

本研究では、移動ロボットによる場所に対応した語彙獲得を目的とし、場所概念の獲得と自己位置情報を修正する方法を検討し、簡易シミュレータおよび、SIGVerse上で実験・評価を行った。場所概念の学習では、多くの場合で目的の場所付近に場所概念がそれぞれ形成されたが、二つの学習対象の場所を一つの場所概念が包含するような場所の分布が形成された場合も見られた。

場所についての知識を獲得することができるロボットは、家の形状や物の位置のような家庭ごとに異なる環境

であっても、各家庭に合った知識を事前に語彙をもつことなく獲得することが可能である。たとえば、オフィスで客人が「どこどこ」へ連れて行ってください」などと言うと、ロボットが目的地まで客人を案内できるようになることが考えられる。また、人がロボットに「ここはどこですか？」と質問すると、ロボットが「ここは“どこどこ”です」と返答を返すことや、場所についての知識と物体について知識を合わせ、「“どこどこ”へ行って“なになに”を取ってきて」などの指示にも対応することが可能になると考えられる。人間とロボットの対話を想定した場合、ロボットが少しの誤りある単語を発話したとしても、人間はその単語が本来なんという単語であるのかを推定し、対話することは可能であると考えられる。

今後の課題としては、場所概念の学習において、自己位置情報を既知と近似していたものを、MCLの自己位置推定情報を使い学習を行うこと、場所概念の数を既知とした学習から未知の状態での学習を行うこと、オフライン学習から逐次的な学習を行えるようにすること、単語のみからの学習ではなく発話文からの語彙獲得を行うことなどが挙げられる。

今後の発展として、間違った場所概念を獲得してしまった場合の場所概念の修正方法や、発話された内容が、今いる場所の名前を表しているのか目的地の名前を表しているのかを判別させる方法なども検討する必要がある。

また本研究は、自己位置推定にMCLを使用したモデルであった。自己位置推定問題において、MCLは最も広く使われる手法であり、それに伴い本手法の適用範囲も広いと考える。しかし、MCLで地図情報を事前に与えることを前提としていたため、地図情報が与えられていないSLAM(Simultaneous Localization And Mapping)[22]を使用したモデルへの拡張も今後の課題である。

参考文献

- [1] 中村, 長井, 岩橋: ロボットによる物体のマルチモーダルカテゴリゼーション; 電子情報通信学会論文誌. D, 情報・システム, Vol. 91, No. 10, pp. 2507–2518 (2008)
- [2] T. Nakamura, T. Araki, T. Nagai and N. Iwahashi: Grounding of word meanings in latent dirichlet allocation-based multimodal concepts; *Advanced Robotics*, Vol. 25, No. 17, pp. 2189–2206 (2011)
- [3] 岩橋: ロボットによる言語獲得: 言語処理の新しいパラダイムを目指して; 人工知能学会誌, Vol. 18, No. 1, pp. 49–58 (2003)
- [4] N. Iwahashi: Robots that learn language: A developmental approach to situated human-robot conversations; (N. Sarkar, editor) *Human Robot Interaction*, InTech (2007)
- [5] N. Iwahashi, R. Taguchi, K. Sugiura, K. Funakoshi and M. Nakano: Robots that learn to converse: Developmental approach to situated language pro- cessing; *Proceedings of International Symposium on Speech and Language Processing*, pp. 532–537 (2009)
- [6] 杉浦, 岩橋, 柏岡, 中村: 言語獲得ロボットによる発話理解確率の推定に基づく物体操作対話; 日本ロボット学会誌, Vol. 28, No. 8, pp. 978–988 (2010)
- [7] J. Hörnstein, L. Gustavsson, J. Santos-Victor and F. Lacerda: Multimodal language acquisition based on motor learning and interaction; *From Motor Learning to Interaction Learning in Robots*, pp. 467–489, Springer (2010)
- [8] D. Roy and A. Pentland: Learning words from sights and sounds: A computational model; *Cognitive Science*, Vol. 26, No. 1, pp. 113–146 (2002)
- [9] P. Gorniak and D. Roy: Probabilistic grounding of situated speech using plan recognition and reference resolution; *Proceedings of the 7th International conference on Multimodal Interfaces*, pp. 138–143, ACM (2005)
- [10] S. Qu and J. Y. Chai: Incorporating temporal and semantic information with eye gaze for automatic word acquisition in multimodal conversational systems; *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 244–253 (2008)
- [11] S. Qu and J. Y. Chai: Context-based word acquisition for situated dialogue in a virtual world; *Journal of Artificial Intelligence Research*, Vol. 37, No. 1, pp. 247–278 (2010)
- [12] P. F. Brown, V. J. Della Pietra, S. A. Della Pietra and R. L. Mercer: The mathematics of statistical machine translation: Parameter estimation; *Computational linguistics*, Vol. 19, No. 2, pp. 263–311 (1993)
- [13] A. S. Huang, S. Tellex, A. Bachrach, T. Kollar, D. Roy and N. Roy: Natural language command of an autonomous micro-air vehicle; *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 2663–2669 (2010)
- [14] 本田, ウォンアラム, 坪山, 長谷川: 半教師ありニューラルネットワークによる場所細胞の獲得とロボットの自律移動制御; 電子情報通信学会論文誌. D, 情報・システム, Vol. 93, No. 4, pp. 469–485 (2010)
- [15] 田口, 岩橋, 船越, 中野, 能勢, 新田: 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)
- [16] 山田, 服部, 田口, 梅崎, 保黒, 岩橋, 船越, 中野: 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)
- [17] 李, 河原: Julius を用いた音声認識インターフェースの作成; ヒューマンインターフェース学会誌, Vol. 11, No. 1, pp. 31–38 (2009)
- [18] 北川: 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)
- [19] 上田, 新井, 浅沼, 梅田, 大隅: パーティクルフィルタを利用した音声認識の実験; 電子情報通信学会論文誌. D, 情報・システム, Vol. 93, No. 4, pp. 469–485 (2010)

- 用した自己位置推定に生じる致命的な推定誤りからの回復法; 日本ロボット学会誌, Vol. 23, No. 4, pp. 466–473 (2005)
- [20] T. Inamura et al.: Simulator platform that enables social interaction simulation –SIGVerse: SocioIntelli-Genesis simulator–; IEEE/SICE International Symposium on System Integration, pp. 212–217 (2010)
- [21] C. M. ビショップ (著), 元田, 栗田, 樋口, 松本, 村田 (監訳): パターン認識と機械学習 上 ベイズ理論による統計的予測, 丸善出版 (2007)
- [22] S. Thrun, W. Burgard, D. Fox, 上田 (訳): 確率ロボティクス, 毎日コミュニケーションズ (2007)

付 錄

付録 1. 場所概念の学習の詳細

場所概念学習アルゴリズムの各ステップについて説明する。

(1) μ, Σ を初期化

初期の場所の分布はすべて, μ_c は任意の範囲内に一様乱数とし, $\Sigma_c = \begin{bmatrix} \sigma_{initial} & 0 \\ 0 & \sigma_{initial} \end{bmatrix}$ とする。

(2) \mathbf{W} の情報なしで, C_t をサンプリング

どのデータがどの場所概念の認識状態を表すかを推定する。 C_{T_o} 以外のすべての変数が与えられたうえでの C_{T_o} の確率を表す式を, (A1) 式に示す。

$$\begin{aligned} p(C_{T_o} | x_{T_o}, O_{T_o}, \mathbf{W}, \mu, \Sigma) \\ = \prod_{t \in T_o} p(C_t | x_t, O_t, \mathbf{W}, \mu, \Sigma) \\ \propto \prod_{t \in T_o} p(x_t | \mu_{C_t}, \Sigma_{C_t}) p(O_t | W_{C_t}) p(C_t) \end{aligned} \quad (\text{A1})$$

これを, (A2) 式のように $t \in T_o$ についてそれぞれサンプリングする。

$$\begin{aligned} C_t \sim p(C_t | x_t, O_t, \mathbf{W}, \mu, \Sigma) \\ \propto p(x_t | \mu_{C_t}, \Sigma_{C_t}) p(O_t | W_{C_t}) p(C_t) \end{aligned} \quad (\text{A2})$$

C_t のサンプリングは, $C_t \in \mathbf{C} = \{1, 2, \dots, L\}$ についてそれぞれ確率を出し, 亂数により選択する。ただし, ステップ(2)では \mathbf{W} の情報が推定されていて, (A3) 式を計算し, サンプリングする。

$$C_t \sim p(x_t | \mu_{C_t}, \Sigma_{C_t}) \quad (\text{A3})$$

(3) μ, Σ をサンプリング

場所の分布の推定の式を, (A4) 式に示す。

$$\begin{aligned} p(\mu, \Sigma | x_{T_o}, O_{T_o}, C_{T_o}, \mathbf{W}) \\ = \prod_{c \in \mathbf{C}} \left(\prod_{\substack{c = C_t \\ t \in T_o}} p(x_t | \mu_{C_t}, \Sigma_{C_t}) \right) \\ \times p(\mu_c, \Sigma_c) \end{aligned} \quad (\text{A4})$$

これを, (A5) 式のように $c \in \mathbf{C} = \{1, 2, \dots, L\}$ につ

いてそれぞれ計算し, サンプリングする。

$$\begin{aligned} \mu_c, \Sigma_c \sim & \left(\prod_{\substack{c = C_t \\ t \in T_o}} p(x_t | \mu_{C_t}, \Sigma_{C_t}) \right) \\ & \times p(\mu_c, \Sigma_c) \end{aligned} \quad (\text{A5})$$

(A5) 式の尤度関数部分は, 教示データ ($t \in T_o$) の中で $C_t = c$ であるデータのみを選び, $p(\mu_c, \Sigma_c)$ に掛け合わせたものである。 (A6) 式のように, 事前分布 $p(\mu_c, \Sigma_c)$ をガウス-ウィシャート分布とし, ガウス分布に対するベイズ推論 [21] を行うことで事後分布を計算する。 A は精度行列であり, 共分散行列の逆行列である ($A = \Sigma^{-1}$)。 m_0, β_0, V_0, ν_0 は, 事前分布のハイパーパラメータである。

$$\begin{aligned} p(\mu_c, A_c) \\ = \mathcal{N}(\mu_c | m_0, (\beta_0 A_c)^{-1}) \mathcal{W}(A_c | V_0, \nu_0) \end{aligned} \quad (\text{A6})$$

μ_c, A_c のサンプリングは, (A7) 式, (A8) 式のように表せる。 N_c は, 教示データの中で $C_t = c$ であるデータの数である。 $m_{N_c}, \beta_{N_c}, V_{N_c}, \nu_{N_c}$ は, 事後分布のハイパーパラメータである。

$$A_c \sim \mathcal{W}(A_c | V_{N_c}, \nu_{N_c}) \quad (\text{A7})$$

$$\mu_c \sim \mathcal{N}(\mu_c | m_{N_c}, (\beta_{N_c} A_c)^{-1}) \quad (\text{A8})$$

教示データ中に該当する場所概念が存在しない場合, パラメータ m_0 を任意の範囲内となる一様乱数により与える。

(4) \mathbf{W} をサンプリング

場所の名前の推定の式を, (A9) 式に示す。

$$\begin{aligned} p(\mathbf{W} | x_{T_o}, O_{T_o}, C_{T_o}, \mu, \Sigma) \\ = \prod_{c \in \mathbf{C}} \left(\prod_{\substack{c = C_t \\ t \in T_o}} p(O_t | W_{C_t}) \right) p(W_c) \end{aligned} \quad (\text{A9})$$

これを, (A10) 式のように $c \in \mathbf{C} = \{1, 2, \dots, L\}$ についてそれぞれ計算し, サンプリングする。

$$W_c \sim \left(\prod_{\substack{c = C_t \\ t \in T_o}} p(O_t | W_{C_t}) \right) p(W_c) \quad (\text{A10})$$

このとき, W_c は, 発話された教示データの言語情報から選ばれるものとし, $p(W_c)$ は, $p(W_c) = 1/N$ のように一様分布として近似する。

W_c のサンプリングは, $t \in T_o = \{t_1, t_2, \dots, t_N\}$ についてそれぞれ確率を出し, 亂数により選択する。教示データの中に該当する場所概念がない場合, $p(W_c)$ が一様分布より, すべての発話された教示データの言語情報から等確率で選択する。

(5) \mathbf{W} の情報ありで, C_t をサンプリング

\mathbf{W} が推定されているため, (A2) 式を計算し, C_t をサンプリングする。サンプリング方法は, ステッ

プ(2)と同様である。

(6) ステップ(3)～(5)の過程を複数回繰り返す。

著者略歴

谷 口

あきら
彰



1990年4月3日生。2013年3月立命館大学情報理工学部知能情報学科卒業。現在、立命館大学大学院情報理工学研究科情報理工学専攻博士前期課程人間情報科学コース在学中。移動ロボットを使用した場所に関する語彙獲得の研究に従事。

吉 崎

はる 陽 紀



1989年2月8日生。2012年3月立命館大学情報理工学部知能情報学科卒業。現在、立命館大学大学院情報理工学研究科情報理工学専攻博士前期課程人間情報科学コース修了。移動ロボットを使用した人の移動の際の地図生成に関する研究に従事。

稻 邑 哲 也



1973年1月23日生。1995年東京大学工学部卒業。1997年度日本学術振興会特別研究員(DC1)。2000年同大学大学院工学系研究科情報工学専攻博士課程修了。同年JST-CREST研究員、2003年東京大学大学院情報理工学系研究科知能機械情報学専攻講師、2006年より国立情報学研究所助教授、総合研究大学院大学複合科学研究科情報学専攻助教授、2007年より同准教授。ヒューマンロボットインタラクション、確率的情報処理を利用したロボットの知能化などの研究に従事。博士(工学)。2003年および2006年人工知能学会全国大会優秀論文賞、2003年船井情報科学奨励賞、2008年日本ロボット学会研究奨励賞受賞、日本人工知能学会、日本機械学会、電子情報通信学会、日本音響学会、IEEE各会員。

谷 口 忠 大 (正会員)



1978年6月24日生。2006年京都大学工学研究科博士課程修了。2005年より日本学術振興会特別研究員(DC2)、2006年より同(PD)。2007年より京都大学情報学研究科にて(PD)再任。2008年より立命館大学情報理工学部助教、2010年より同准教授。個体と組織における記号過程の計算論的な理解や共生社会に向けた知能情報学技術の応用研究についての研究に従事。京都大学博士(工学)。計測自動制御学会学術奨励賞、システム制御情報学会学会賞奨励賞、論文賞、砂原賞など受賞。計測自動制御学会、日本人工知能学会、日本神経回路学会、IEEEなどの会員。