2D1-OS-12a-3

実ロボットを用いた自己位置と語彙の同時推定による音声言語獲得

Word Acquisition from Speech by Simultaneous Estimation of Self-Location and Vocabulary on a Real Robot

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

We have proposed a method which enables a robot to acquire words related to places from self-locations and uttered sentences. In previous experiments, in contrast with that uttered sentences were real speech data experiments using a mobile robot were performed in a simulator environment. In this paper, we perform an experiment about lexical acquisition using a real robot. The experimental result showed that our proposed method enables a robot to obtain words related to places appropriately even in the real environment.

1. はじめに

屋内で動作するサービスロボットなどを想定した場合，ロボットは多様な人間の生活環境において，自らが持つセンサから得られる情報をもとに周囲の状況を認識する必要がある [Thrun 05]．さらに，人との言語によるインタラクションを通して環境中の特定の領域に対して人が割り当てた単語とその語が指し示す領域を獲得することが重要である．本研究では，事前に単語の知識を持たず日本語の音節のみ認識可能な自律移動ロボットに，その場その場で発話文により教示を行うことで，場所に関する語彙を獲得させることを目的とする．

以上の目的の下，我々は自己位置と語彙の同時推定モデル [谷口 14a] および，その拡張であるノンパラメトリックベイズ法により自己位置推定に発話文の教師なし形態素解析結果を統合した場所概念獲得モデル [谷口 14b] を提案している．しかしこれまでの実験では，発話文には実際の音声を使用していたものの，シミュレータ上に構築した環境での小語彙の実験しか行えていなかった．そこで本稿では，実ロボットを用いた場所概念の獲得に関する実験を行い，実環境上における本手法の有効性について検証する．

図 1: 場所概念獲得の概要図

2. 先行研究

語彙を持たないロボットが，多様な言い回しを含む発話から単語の正しい分節，音素系列，単語と対象間の関係を学習する手法 [田口 10] やその拡張である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 [田口 11]．これらの研究では，学習した語彙知識をロボット自身の自己位置推定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を含んだ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ことが可能である．

3. 場所概念獲得モデル [谷口 14b]

場所概念獲得モデルは，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 Localization) [Thrun 05] に場所概念を導入した確率的生成モデルである．

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

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，位置の空間的な広がりを位置分布と呼ぶ．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獲得，(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について問題とする．本研究の全体像を表す概略図を図 1 に示す．

3.1 モデルの概要

場所概念獲得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．

提案手法の全体的な特徴三つを以下に示す．(i) ノンパラメトリックベイズ法を導入することによりデータに応じて適切な場所概念の数 L，位置分布の数 K を学習できる．Dirichlet Process の構成法としては，SBP (Stick Breaking Process) [Sethuraman 94] を用いる．(ii) 同一の場所の名前に対し異なる複数の位置分布を対応させることができる．つまり，多項分

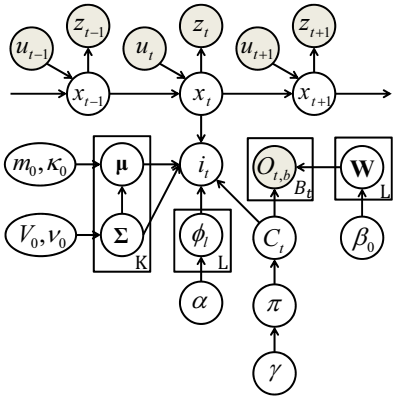

図 2: 場所概念獲得モデルのグラフィカルモデル

表 1: グラフィカルモデルの要素表

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
i_t	位置分布の index
$O_{t,b}$	時刻 t における b 番目の分割単語
π	場所概念の index の多項分布
ϕ_l	位置分布の index の多項分布
$\mathbf{W}$	場所の名前の多項分布
$\boldsymbol{\mu}, \boldsymbol{\Sigma}$	位置分布（平均，共分散行列）
α,γ,β_0, m_0,κ_0, V_0,ν_0	各ハイパーパラメータ

布で表した一つの場所の名前の確率分布に対し，混合ガウス分布が対応可能となる．(iii) 音節認識ラティスの教師なし形態素解析手法 [Neubig 12] により，発話文から音節認識のゆらぎを抑えた単語分割結果を得ることができる．

3.2 場所概念の学習

場所概念の学習は，複数回の教示データを溜め込んだ後，オフラインで学習を行う．このとき，教示された時刻 t の集合は $T_o = \{t_1, t_2, ..., t_N\}$ である．N は教示データ数である．時刻ごとの位置情報や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 $\Theta = \{\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, \pi\}$ をギブスサンプリングによって推定する．教示の際は，自己位置推定を行っているロボットに，教示対象場所で文章による発話を複数回行う．発話文は，教師なし形態素解析器により単語に分割され，単語集合 $O_{t,\mathbf{B}}$ として BoW (Bag-of-Words) 形式で与える．

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．詳細については [谷口 14b] を参照されたい．位置分布の index i_t の事後分布を (1) 式に示す．

$$p(i_t = k \mid x_t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, C_t) \propto \mathcal{N}(x_t \mid \mu_{k=i_t}, \Sigma_{k=i_t}) \mathrm{Mult}(i_t = k \mid \phi_{l=C_t}) \quad (1)$$

場所概念の index C_t の事後分布を (2) 式に示す．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の集合である．

図 3: 自律移動ロボット：Turtlebot2

$$p(C_t = l \mid x_t, i_t, O_{t,\mathbf{B}}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi_l, \pi) \propto \mathrm{Mult}(O_{t,\mathbf{B}} \mid W_{l=C_t}) \mathrm{Mult}(i_t = k \mid \phi_{l=C_t}) \times \mathrm{Mult}(C_t = l \mid \pi) \quad (2)$$

場所の名前 $\mathbf{W}$ は，$l \in L$ ごとに (3) 式のようにサンプリングされる．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 $C_t = l$ である発話文の集合である．

$$\mathrm{Dir}(W_l \mid \beta_{n_l}) \propto \mathrm{Mult}(\mathbf{O}_l \mid W_l) \mathrm{Dir}(W_l \mid \beta_0) \quad (3)$$

位置分布 $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ は，$k \in K$ ごとに (4) 式のようにサンプリングされる．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 $i_t = k$ である位置の集合である．$\mathcal{N}$-$\mathcal{W}(\cdot)$ はガウス-ウィシャート分布である．

$$\mathcal{N}\text{-}\mathcal{W}(\mu_k, \Sigma_k \mid m_{n_k}, \kappa_{n_k}, V_{n_k}, \nu_{n_k}) \propto \mathcal{N}(\mathbf{x}_k \mid \mu_k, \Sigma_k) \mathcal{N}\text{-}\mathcal{W}(\mu_k, \Sigma_k \mid m_0, \kappa_0, V_0, \nu_0) \quad (4)$$

場所概念の index の多項分布 π に関する事後分布を (5) 式に示す．

$$\mathrm{Dir}(\pi \mid C_{T_o}, \gamma) \propto \mathrm{Mult}(C_{T_o} \mid \pi) \mathrm{Dir}(\pi \mid \gamma) \quad (5)$$

位置分布の index の多項分布 ϕ_l は，$l \in L$ ごとに (6) 式のようにサンプリングされる．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 の集合である．

$$\mathrm{Dir}(\phi_l \mid \mathbf{i}_l, \alpha) \propto \mathrm{Mult}(\mathbf{i}_l \mid \phi_l) \mathrm{Dir}(\phi_l \mid \alpha) \quad (6)$$

4. 実験

自律移動ロボット Turtlebot2[*1] を使用し，実環境上での場所概念の学習に関する実験を行う．

4.1 実験条件

実験に使用した Turtlebot2 を図 3 に示す．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．ロボット前方には距離センサとして Kinect が搭載されている．音声認識器には，大語彙連続音声認識システム Julius[*2] を使用した．Julius の単語辞書には，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 社製の PG27-USB を使用した．形態素

*1 http://turtlebot.com/
*2 使用バージョン：dictation-kit-v4.3.1-linux GMM 版，http://julius.sourceforge.jp/

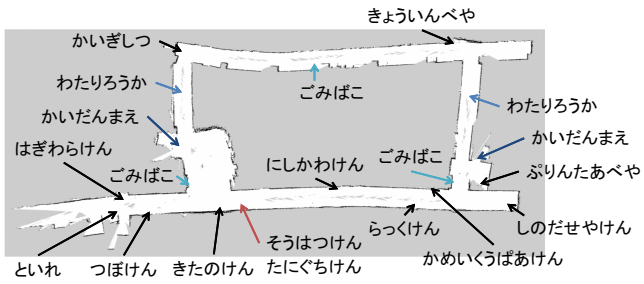

図 4: 地図と教示場所と教示した場所の名前

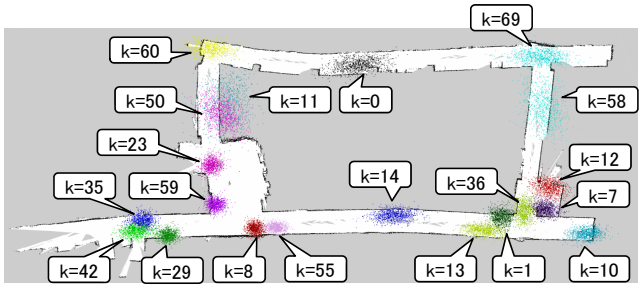

図 5: 各位置分布の学習結果

表 2: 発話文の言い回しの種類

** 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

解析器には，教師なし形態素解析手法 [Neubig 12] が実装された latticelm[*3]を使用した．

本実験では，SLAM (Simultaneous Localization And Mapping) [Thrun 05] により事前に地図生成を行った後，生成した地図を用いて自己位置推定を行う．地図生成および自己位置推定には，ROS (Robot Operating System) の Hydro 上のパッケージを使用した．地図は 2 次元の占有格子地図である．自己位置推定の際，初期パーティクルはロボットの初期位置とした．パーティクル数は $M = 2000$ である．x_t については教示した際の自己位置推定結果を用いるものとする．本実験では，位置情報の取得と音声の取得は別々に行った．各モデルパラメータ値は，$L = 100$，$K = 100$，$\alpha = 10$，$\gamma = 20$，$\beta_0 = 0.2$，$m_0 = [0, 0]^{\mathrm{T}}$, $\kappa_0 = 0.001$, $V_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\nu_0 = 2$ とし，イテレーション回数は，100 回とした．

実験を行う環境上で生成した地図および各場所の名前に対する発話教示場所を図 4 に示す．教示した場所の数は 19，教示した場所の名前の語彙数は 17 である．それぞれの教示した場所の名前に対し 5 回ずつ，合計 100 回分の発話教示を行った．青い矢印で示した箇所は，異なる場所において同一な名前のある箇所を示している．赤い矢印で示した箇所は，同一な場所において異なる名前のある箇所を示している．発話の際に使用した文章の言い回しを表 2 に示す．

4.2 実験結果

場所概念の学習結果の 1 例を以下に示す．地図上に学習された位置分布を図示したものを図 5 に示す．各色の点群は，学習された位置分布に従う点を 1000 個ずつ描画したものである．各位置分布の色はランダムに決定した．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．表 3，4 は，場所概念の index $C_t = 10$ のときの場所の名前 W_{10} と位置分布の index の多項分布 ϕ_{10} の要素を各確率値の降順に並べ，上位 5 個のみを示したものである．この結果より，“とりれ” という単語が $k = 42$ の位置分布と対応して学習されていることがわかる．同様に，表 5，6 では “きだのけん” が $k = 8$ の位置分布と対応して学習されており，表 7，8 では “かいぎひつ” が $k = 60$ の位置分布と対応して学習されていることがわかる．また，表 9，10 では “しのざせや”，“けん” のように二つの単語に分割された状態で $k = 10$ の位置分布と対応して学習されている．表 11，12 からは，$C_t = 27$ の場所概念が $k = 55$ の位置分布を示していることがわかる．この結果より，“そはつけ”，“ん” や “たに”，“ぐち” などのように教示した場所の名前に対し細かく単語分割されているものの，同一の場所に対し複数の名称がある場合でも学習可能であることが示された．表 13，14 では，二番目ではあるが “ごみばこ” という単語が高い確率値を示しており，三つの教示場所に構成された位置分布 $k = 0, 36, 59$ が高い確率値を示している．この結果より，複数の異なる場所に対し，同一の名称が割り当てられている場合でも学習可能であることが示された．表 15，16 では，別々の場所概念として学習されることを期待したが “わたりろおか” と “かいだんまえ” が同じ $C_t = 2$ の場所概念として学習された．そのため，ϕ_2 も双方に対応する教示場所に構成された位置分布の index k が高い確率値を示す結果となった．

表 3: 場所の名前 W_{10} の確率の高い単語

単語	確率
とりれ	**0.1543**
わ	0.1176
ここが	0.0799
にきました	0.0452
びゃよ	0.0431

表 4: 位置分布の index の多項分布 ϕ_{10} の確率の降順

位置分布の index k	確率
42	**0.3400**
83	0.0077
9	0.0075
46	0.0075
11	0.0075

表 5: 場所の名前 W_{29} の確率の高い単語

単語	確率
きだのけん	**0.2087**
です	0.0909
だよ	0.0900
あ	0.0497
このばしょが	0.0496

表 6: 位置分布の index の多項分布 ϕ_{29} の確率の降順

位置分布の index k	確率
8	**0.3356**
60	0.0087
70	0.0080
17	0.0080
2	0.0079

*3 使用バージョン：latticelm 0.4,
http://www.phontron.com/latticelm/

表 7: 場所の名前 W_{32} の確率の高い単語

単語	確率
かいぎひつ	**0.1812**
にきました	0.1128
ごみらこ	0.0788
あ	0.0780
だよ	0.0756

表 8: 位置分布の index の多項分布 ϕ_{32} の確率の降順

位置分布の index k	確率
60	**0.3013**
0	0.0653
7	0.0637
36	0.0072
33	0.0069

表 9: 場所の名前 W_6 の確率の高い単語

単語	確率
しのざせや	**0.1775**
けん	**0.1739**
です	0.0745
ここ	0.0417
わ	0.0412

表 10: 位置分布の index の多項分布 ϕ_6 の確率の降順

位置分布の index k	確率
10	**0.3387**
37	0.0081
5	0.0080
94	0.0078
29	0.0076

5. おわりに

本稿では，実環境での実ロボットによる場所概念獲得モデルの有効性の検証を行った．場所概念の学習の実験により，実環境においても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ることを確認した．[Neubig 12] の教師なし形態素解析については，[谷口 14b] で得られた結果と同様，教示した発話全体に対して音節認識のゆらぎを抑える効果が見られたが，場所の名前が細かく単語分割される場合も見られた．

誤りを含んだ音声認識結果から教師なしで高精度な語彙獲得を行う研究として，物体概念と言語モデルを相互に学習させる手法 [Nakamura 14] や音素認識結果と教師なし形態素解析による言語モデルを交互に繰り返し更新する手法 [Heymann 14] が提案されている．本稿では，場所概念獲得モデル [谷口 14b] を利用した実環境実験であったが，さらに精度の高い語彙獲得のために，[Nakamura 14, Heymann 14] と似たアプローチである場所概念と言語モデルの相互推定手法 [谷口 14c] を適用することにより，実環境においても高精度な場所に関する語彙獲得が可能になることが期待される．

表 11: 場所の名前 W_{27} の確率の高い単語

単語	確率
たに	**0.0981**
そはつけ	**0.0980**
ん	**0.0978**
ぐち	**0.0961**
です	0.0784

表 12: 位置分布の index の多項分布 ϕ_{27} の確率の降順

位置分布の index k	確率
55	**0.5065**
84	0.0062
21	0.0061
42	0.0061
37	0.0059

表 13: 場所の名前 W_8 の確率の高い単語

単語	確率
です	0.1949
ごみばこ	**0.1804**
ここ	0.1019
あ	0.0813
らぷけん	0.0425

表 14: 位置分布の index の多項分布 ϕ_8 の確率の降順

位置分布の index k	確率
36	**0.1742**
59	**0.1357**
0	**0.1356**
13	0.1353
9	0.0055

表 15: 場所の名前 W_2 の確率の高い単語

単語	確率
わたりろおか	**0.1654**
かいだんまえ	**0.1509**
です	0.1170
にきました	0.0687
えわ	0.0683

表 16: 位置分布の index の多項分布 ϕ_2 の確率の降順

位置分布の index k	確率
23	**0.1746**
58	**0.1741**
50	**0.1424**
12	**0.1408**
11	**0.0386**

参考文献

[Heymann 14] Heymann, J., Walter, O., Haeb-Umbach, R., and Raj, B.: Iterative Bayesian Word Segmentation for Unspuervised Vocabulary Discovery from Phoneme Lattices, in *39th International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)* (2014)

[Nakamura 14] Nakamura, T., Nagai, T., Funakoshi, K., Nagasaka, S., Taniguchi, T., and Iwahashi, N.: Mutual learning of an object concept and language model based on MLDA and NPYLM, in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*, pp. 600–607 (2014)

[Neubig 12] Neubig, G., Mimura, M., and Kawahara, T.: Bayesian learning of a language model from continuous speech, *IEICE TRANSACTIONS on Information and Systems*, Vol. 95, No. 2, pp. 614–625 (2012)

[Sethuraman 94] Sethuraman, J.: A CONSTRUCTIVE DEFINITION OF DIRICHLET PRIORS, *Statistica Sinica*, Vol. 4, pp. 639–650 (1994)

[Thrun 05] Thrun, S., Burgard, W., and Fox, D.: *Probabilistic Robotics*, MIT Press (2005)

[谷口 14a] 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)

[谷口 14b] 谷口彰, 稲邑哲也, 谷口忠大 : 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, 人工知能学会全国大会論文集, 第 28 巻 (2014)

[谷口 14c] 谷口彰, 稲邑哲也, 谷口忠大 : 場所概念と言語モデルの相互推定によるロボットの場所に関する語彙獲得, 日本ロボット学会学術講演会, 第 32 巻 (2014)

[田口 10] 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)

[田口 11] 田口亮, 山田雄治, 服部公央亮, 梅崎太造, 保黒政大, 岩橋直人, 船越孝太郎, 中野幹生 : 連続音声と自己位置から場所の名前を学習するロボット, 人工知能学会全国大会論文集, 第 25 巻 (2011)