4D2-OS-37d-5

マルチモーダル情報に基づく階層的場所概念形成

Formation of Hierarchical Spatial Concept based on Multi-modal Information

萩原良信 *1　井上将一 *1　谷口忠大 *1
Yoshinobu Hagiwara　Masakazu Inoue　Tadahiro Taniguchi

*1立命館大学　情報理工学部
Information Science and Engineering, Ritsumeikan University

In this research, we propose a method to estimate the hierarchical structure of spatial concepts and clarifies the influence and effect of the proposed method. The hierarchical spatial concept is formed by hierarchical multimodal LDA based on multi-modal information: the position, visual, language information. The position and visual information are acquired by MCL and CNN, respectively. The language information is acquired as Bag of Words. Experimental results demonstrate the effectiveness of the proposed method in the scene of human-robot communication at home environments.

1. はじめに

事物のカテゴリ分類は人の認知機能にとって重要であることが指摘されている [Ashby 05]．また，人がマルチモーダル情報を用いて分類したカテゴリは，人が事物を理解するための概念と関係している．人と共存するロボットは，人の持つ概念に基づく物体や場所の言語表現をマルチモーダルな情報に基づいて理解する必要がある．

先行研究として，観測情報に基づいて物体や場所の概念をロボットがボトムアップに形成する手法が提案されている．Nakamura らは，自然言語処理の分野で研究されてきた統計モデルの Latent Dirichlet Allocation (LDA) を拡張し，ロボットに物体概念を教師なしで獲得させる手法である Multimodal LDA を提案している [Nakamura 09]．ここで，物体概念とは，ロボット自身が得ることのできるマルチモーダル (視覚，触覚，聴覚) 情報から形成した物体のカテゴリのことを意味している．Ando らは，LDA に nested Chinese restaurant process(nCRP) を導入した hierarchical latent Dirichlet allocation (hLDA) を，マルチモーダル情報の分類が可能なモデルである，Hierarchical Multimodal Latent Dirichlet Allocation (hMLDA) に拡張することによって，物体の階層的なカテゴリ分類を学習する手法を提案している [Ando 13]．Taniguchi らは，音声認識結果と自己位置推定結果からロボットが場所概念を獲得する手法を提案している [Taniguchi 14]．ここで，場所概念とは，場所の名前と，場所の分布によって定義されている．ロボットは，場所の名前を人の発話から観測する事で場所概念を形成する．Ishibushi らは，画像認識結果と自己位置推定結果から，ロボットが場所の領域を学習する手法を提案している [Ishibushi 15]．

しかしながら，これらの先行研究では，場所に潜む階層構造については議論されて来なかった．実世界における場所は，それぞれの場所が独立しているわけではなく, 階層的な構造を持っている. さらに人間同士の言語的コミュニケーションでは，場所の階層構造を考慮することによって, より円滑な意思疎通が行われている．例えば, 一般的な家庭環境を想定し「リビング」「ダイニング」「テーブル前」等の場所が存在するときに，人は「ダイニングのテーブル前」「リビングのテーブル前」のような場所の階層構造を考慮した表現を用いることで意図する場所を明確に表現することができる．ここで「ダイニング」と「テーブル前」の関係は part-of 関係であり,「テーブル前」という場所は「ダイニング」という場所の一部であると表現できる．part-of 関係とは階層構造の全体，部分の関係を表現するものであり，人は普段から場所の part-of 関係を考慮した表現を用いて会話を行っていると考えられる．

そこで本研究では，ロボットが取得したマルチモーダル情報から階層的な場所概念を形成する手法を提案し，ロボットを用いた実環境における学習実験から提案手法の有用性を評価する．

連絡先: 萩原良信，立命館大学，滋賀県草津市野路東 1 丁目 1-1，077-561-5745，077-561-5745，yhagiwara@em.ci.ritsumei.ac.jp

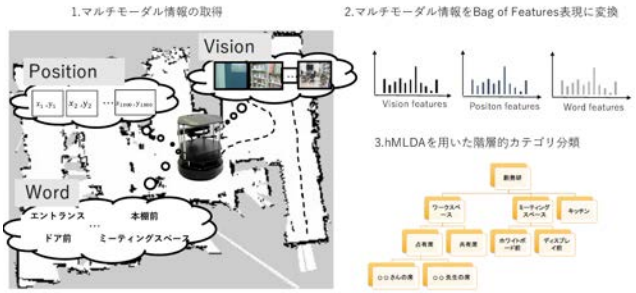

図 1: 提案手法の概要

2. 提案手法

2.1 概要

提案手法の概要を図 1 に示す．まず，ロボットは，Simultaneous Localization and Mapping(SLAM) によって生成した地図に基づいて環境中を移動し，画像情報，自己位置情報，語彙情報を取得する．次に，取得したマルチモーダル情報を Bag of features，Bag of Words に変換し，これを各モダリティの特徴量として扱う．この特徴量を観測情報として，階層的カテゴリ分類を行う．

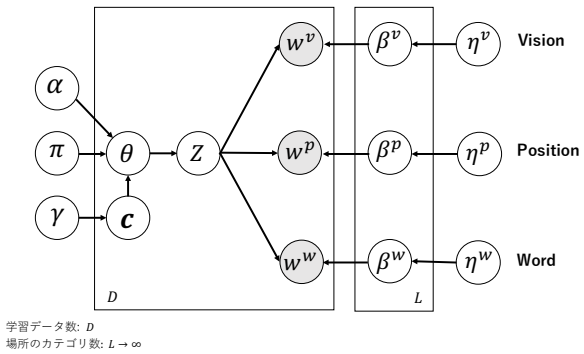

図 2: 提案手法のグラフィカルモデル

表 1: グラフィカルモデルの変数の定義

z	場所のカテゴリ
w^v, w^p, w^w	画像情報，位置情報，語彙情報
$\beta^v, \beta^p, \beta^w$	画像情報，位置情報，語彙情報の多項分布
θ	場所のカテゴリの多項分布
$\boldsymbol{c}$	木構造のパス
η^v, η^v, η^w	ディリクレ事前分布のハイパーパラメータ
γ	$\boldsymbol{c}$ のハイパーパラメータ
α, π	θ のハイパーパラメータ

2.2 マルチモーダル情報の取得

画像情報は，CNN(Convolutional neural network) のフレームワークである Caffe を用いた画像の物体認識の結果を利用する．Caffe の出力結果は，要素数を I としたときの物体 $O(O \in o_1, o_2, ..., o_I)$ の確率分布 $p(o_i)$ として与えられ，モダリティの特徴量を $\boldsymbol{w}^v = [p(o_i) * 10^2]$ とする．

位置情報は，MCL によって推定された地図上の位置座標 (x, y) を用いる．得られた位置座標 x, y に対して，$k = 2$ としたときの k-means 法によってクラスタリングを行い，その結果の各クラスタに対してさらに階層的クラスタリングを行い 64 次元の特徴量 $\boldsymbol{w}^p$ に変換する．

語彙情報は，人が発話した単語をテキストデータに変換したものを用いる．与える語彙の種類は，あらかじめ辞書データとして用意されているものとする．各データに対して与えられた単語と辞書を用いて，Bag of Words 表現の特徴量 $\boldsymbol{w}^w$ を作成する．

2.3 提案モデルの生成過程

提案手法のグラフィカルモデル，変数の定義をそれぞれ図 2，表 1 に示す．図 2 において，$\boldsymbol{c}$ は，γ をパラメータとする nCRP により生成される木構造のパスである．また，z は，場所のカテゴリを表しており，α と π をパラメータとする Stick breaking process によって生成される．w^v, w^p, w^w は，それぞれ画像，位置，語彙情報であり，β^* をパラメータとする多項分布から生成される．β^* は，η^* をパラメータとするディリクレ事前分布に従い決定される．

提案モデルの生成過程について述べる．各モダリティ $(m \in v, p, w)$ において，テーブル $(k \in T)$ で特徴量を生成する確率を表す多項分布のパラメータ β_k^m を決める．ここで，T はテーブルの集合である．

$$\beta_k^m \sim Dirichlet(\eta^m) \tag{1}$$

各データ $(d \in 1, 2, ..., D)$ に対して以下の操作を繰り返す．

- c_d をデータ d に対する木構造のパスとし，γ をパラメータとした nCRP により c_d を決定．

$$\boldsymbol{c}_d \sim nCRP(\gamma) \tag{2}$$

- Stick breaking process により，多項分布のパラメータ θ を生成．

$$\theta_d \sim GEM(\alpha, \pi) \tag{3}$$

- データ d におけるすべてのモダリティ m の n 番めの特徴量に対して以下を繰り返す．
  - 特徴量 $w_{d,n}^m$ のカテゴリ $z_{d,n}^m$ を決定

$$z_{d,n}^m \sim Multi(\theta_d) \tag{4}$$

  - パス c_d のカテゴリ $z_{d,n}^m$ から特徴量 $w_{d,n}^m$ を生成

$$w_{d,n}^m \sim Multi(\beta_{\boldsymbol{c}_d}[z_{d,n}^m]) \tag{5}$$

    ここで，$\beta_{\boldsymbol{c}_d}[z_{d,n}^m]$ は，データ d のパス $\boldsymbol{c}_d$ におけるカテゴリ $z_{d,n}^m$ に対するパラメータ β を表している．

2.4 提案モデルのパラメータの学習

パラメータの学習は，Gibbs sampling により行う．データごとのパス c_d と，そのパスにおけるデータ d のモダリティ m の n 番目の特徴量に割り当てられるカテゴリ $z_{d,n}^m$ を交互にサンプリングすることによってパラメータを学習する．

カテゴリ $z_{d,n}^m$ は以下に従ってサンプリングされる．

$$\begin{aligned} z_{d,n}^m &\sim p(z_{d,n}^m | \boldsymbol{z}_{-(d,n)}^m, \boldsymbol{c}, \boldsymbol{w}^m, \alpha, \pi, \eta^m) \\ &\propto p(z_{d,n}^m | \boldsymbol{z}_{d,-n}^m, \alpha, \pi) p(w_{d,n}^m | \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{w}_{-(d,n)}^m, \eta^m) \end{aligned} \tag{6}$$

式 (6) において $p(z_{d,n}^m | \boldsymbol{z}_{d,-n}^m, \alpha, \pi)$ は，Stick breaking process によって生成される多項分布であり，以下の確率によって求められる．

$$\begin{aligned} &p(z_{d,n}^m | \boldsymbol{z}_{d,-n}^m, \alpha, \pi) \\ &= E\left[V_k \prod_{j=1}^{k-1} (1 - V_j) | \boldsymbol{z}_{d,-n}^m, \alpha, \pi \right] \\ &= \frac{(1-\alpha)\pi + \#[\boldsymbol{z}_{d,-n}^m = k]}{\pi + \#[\boldsymbol{z}_{d,-n}^m \geq k]} \prod_{j=1}^{k-1} \frac{\alpha\pi + \#[\boldsymbol{z}_{d,-n}^m > j]}{\pi + \#[\boldsymbol{z}_{d,-n}^m \geq j]} \end{aligned} \tag{7}$$

ここで $\#[\cdot]$ は，与えられた条件を満たす数であり，V_j は Stick breaking process におけるパラメータである．また，式 (6) において $p(w_{d,n}^m | \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{w}_{-(d,n)}^m, \eta^m)$ は，パス $\boldsymbol{c_d}$，カテゴリ $z_{d,n}^m$ から特徴量が生成される確率であり，特徴量を生成する多項分布のパラメータが，ディリクレ事前分布に従って生成されると仮定しているため，以下の式が得られる．

$$\begin{aligned} &p(w_{d,n}^m | \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{w}_{d,n}^m, \eta^m) \\ &\propto \#[\boldsymbol{z}_{-(d,n)}^m = z_{d,n}^m, \boldsymbol{c}_{z_{d,n}^m} = c_{d,z_{d,n}^m}, \boldsymbol{w}_{-(d,n)}^m = w_{d,n}^m] \\ &\quad + \eta^m \end{aligned} \tag{8}$$

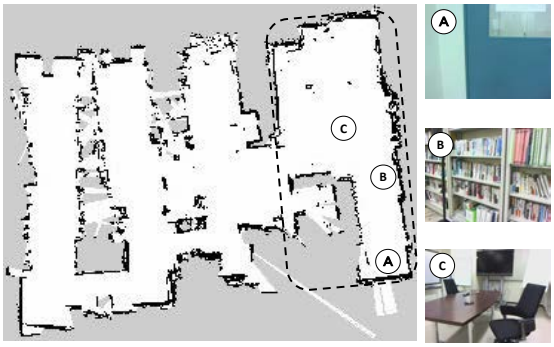

図 3: 実験環境の地図と撮像画像

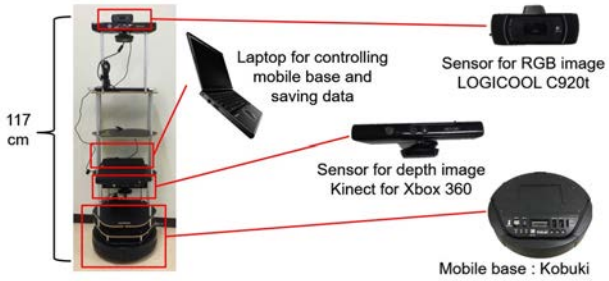

図 4: 実験に用いたロボットの構成

これはパス $\boldsymbol{c}_d$ において，特徴量 $w^m_{d,n}$ にカテゴリ $z^m_{d,n}$ が割り当てられた回数を表す．

パス c_d のサンプリングは以下に従ってサンプリンされる．

$$\begin{aligned}\boldsymbol{c}_d \sim p(\boldsymbol{c}_d|\boldsymbol{w}^v,\boldsymbol{w}^p,\boldsymbol{w}^w,\boldsymbol{c}_{-d},\boldsymbol{z},\eta^v,\eta^p,\eta^w,\gamma)\\ \propto p(\boldsymbol{c}_d|\boldsymbol{c}_{-d},\gamma)\\ \times p(\boldsymbol{w}^v_d|\boldsymbol{c},\boldsymbol{w}^v_{-d},\boldsymbol{z}^v,\eta^v)p(\boldsymbol{w}^p_d|\boldsymbol{c},\boldsymbol{w}^p_{-d},\boldsymbol{z}^p,\eta^p)\\ \times p(\boldsymbol{w}^w_d|\boldsymbol{c},\boldsymbol{w}^w_{-d},\boldsymbol{z}^w,\eta^w)\end{aligned} \tag{9}$$

ここで $\boldsymbol{c}_{-d}$ は $\boldsymbol{c}$ から $\boldsymbol{c_d}$ を除いたパスの集合を表している．

各学習データ $d \sim 1,...,D$ に対して式 (8)，式 (9) に基づくサンプリングを繰り返し行うことで，全学習データのパスとカテゴリは，$\hat{\boldsymbol{c}}$, $\hat{\boldsymbol{z}}$ へと収束する．

3. 実験

3.1 実験目的

提案手法の有用性を評価するため，ロボットを用いた実環境における階層的な場所概念の形成実験を実施した．

3.2 実験条件

図 3 に実験環境においてロボットが作成した地図と撮像画像を示す．また，実験に用いたロボットの構成を図 4 に示す．カメラの設置位置は，人間の身体を考慮して 117cm とした．これは，6 歳の男子の平均身長に相当する．ロボットは，環境内を移動しながら MCL のパーティクルのリサンプリングのタイミングで自己位置情報と画像情報を観測データとして同時に取得した．本実験では，地図中の破線内で得られた 900 データを学習データとして用いる．

表 2: ワード辞書

創発研	エントランス	ミーティングスペース
扉前	傘立て前	マガジンラック前
本棚前	イス置き場前	スカイプ PC 前
double 前	キーボード前	テーブル前
物置棚前	ディスプレイ前	ホワイトボード前

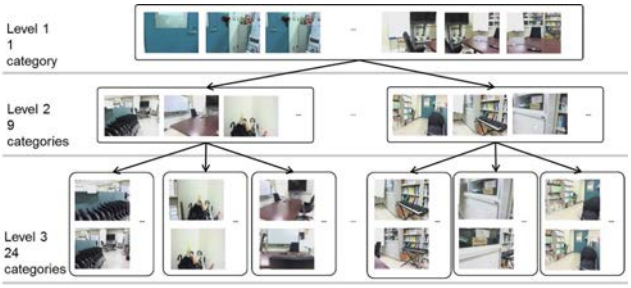

図 5: 形成された階層的場所概念によるカテゴリ分類の結果

また，学習データに語彙を与える際に用いた辞書を表 2 に示す．学習データに与える語彙は，対象となるデータに対して考え得る場所の呼び方を辞書から選択して与えた．一つの位置に対して,「扉前」といった語彙だけでなく「エントランス」といったその空間を表現する語彙や，実験環境全体を意味する「創発研」といった語彙も同時に与えている．このとき，語彙を与えた割合は全学習データの 20% であった．

階層数は 3 とし，モデルのハイパーパラメータである α,π,γ,η は，それぞれ $\alpha = 0.5,\pi = 100,\gamma = 1.0,\eta^v = 9.0\times10^{-2},\eta^p = 3.0\times10^{-4},\eta^w = 3.0\times10^{-6}$ とした．

3.3 実験結果

図 5 に提案手法により形成された階層的場所概念によるカテゴリ分類の結果を示す．Level1 が最上位の階層，Level3 が最下位の階層である．Levle1 のカテゴリ数は 1，Level2 のカテゴリ数は 9，Level3 のカテゴリ数は 24 となった．図 6 は，各カテゴリに属する学習データの座標を地図上にプロットしたものである．また，図 7, 8, 9 は，それぞれ Level1，Level2，Level3 における，各カテゴリから生起される語彙の確率分布である．

Level1 には，学習データ中の全データが割り当てられている category1 のみが存在する．category1 では，実験環境全体を意味する単語である「創発研」が高い確率で生起していることが分かる．

Level2 における category2 では，実験環境内における入り口付近が 1 つのカテゴリとして分類されていることが図 6 から分かる．また，語彙の分布を見ると「エントランス」の生起確率が高くなっている．category8 では，実験環境内におけるエントランスより奥の部分が 1 つのカテゴリとして分類されており「ミーティングスペース」の語彙の生起確率が高くなっている．

Level3 では，Level2 における category2 や category8 に分類されたデータを，さらに細かく分類した結果が得られた．category2 の下位に形成された category11，category12 では，「扉前」「イス置き場前」といった語彙の生起確率が高くなって

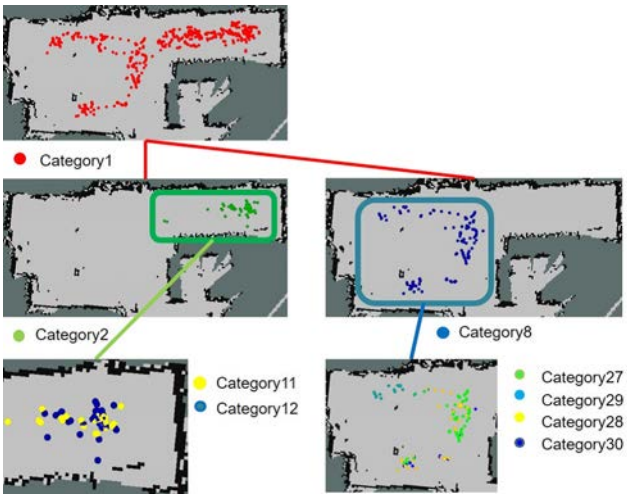

図 6: 階層的場所概念の各カテゴリに属するデータの座標位置

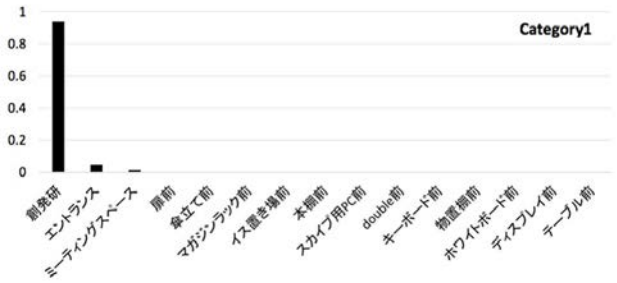

図 7: Level1 のカテゴリから生起される語彙の確率分布

いることが分かる．また，図 5 の category11 や category12 では，実際に「扉前」「イス置き場前」の画像が割り当てられている．例えば，category12 は，level3:「イス置き場前」，level2:「エントランス」，level1:「創発研」といった階層的な解釈が可能なカテゴリ分類ができていることが分かる．

4. おわりに

本稿では，画像，位置, 語彙のマルチモーダル情報に基づいてロボットがボトムアップに階層的な場所概念を形成する手法を提案した．実環境でのロボットを用いた階層的場所概念の学習実験を実施し，提案手法の有用性を評価した．実験結果として，形成された場所概念の語彙の生起確率を異なる階層におい

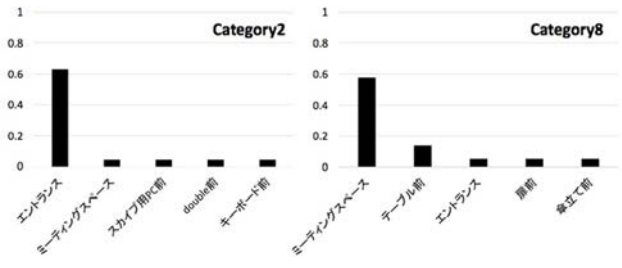

図 8: Level2 の各カテゴリから生起される語彙の確率分布

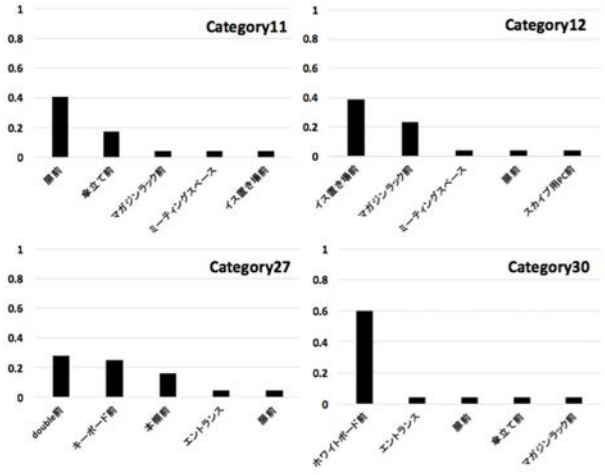

図 9: Level3 の各カテゴリから生起される語彙の確率分布

て出力し，level3:「イス置き場前」，level2:「エントランス」，level1:「創発研」といった，階層構造を考慮した場所概念の形成が確認された．さらに，形成された階層的な場所概念に対して, 実際に人がもつ場所の概念との類似性を検証する実験を実施した．この結果については，紙面の都合のため発表において報告する．今後は，人とロボットの言語的コミュニケーションに提案手法を展開し，マルチモーダル情報に基づく階層的場所概念の有用性について検証していきたい．

参考文献

[Ashby 05] F. Gregory Ashby and W. Todd Maddox : ”human category learning”. Annu. Rev. Psychol., Vol. 56, pp. 149-178, 2005.

[Nakamura 09] T. Nakamura, T. Nagai and N. Iwahashi : “Grounding of word meanings in multimodal concepts using LDA”. IEEE/RSJ International Conference on Intelligent Robots and Systems,pp.39433948, 2009.

[Ando 13] Y. Ando, T. Nakamura, and T Nagai : ”Formation of hierarchical object concept using hierarchical latent dirichlet allocation”. 2013 IEEE/RSJ International Confer ence on Intelligent Robots and Systems (IROS), 2272-2279.

[Taniguchi 14] A. Taniguchi, H. Yoshizaki, T. Inamura, and T. Taniguchi : ”Research on simultaneous estimation of self-location and location concepts”. Transactions of the Institute of Systems, Control and Information Engineers, Vol. 27cle, No. 4, pp. 166-177, 2014.

[Ishibushi 15] Satoshi Ishibushi, Akira Taniguchi, Toshiaki Takano, Yoshinobu Hagiwara and Tadahiro Taniguchi : ”Statistical Localization Exploiting Convolutional Neural Network for an Autonomous Vehicle”, 41th Annual Conference of the IEEE Industrial Electronics Society (IECON), pp1369-1375, 2015.