The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

場所概念形成に基づいた空間のSemantic Mapping

Spatial Concept Formation-based Semantic Mapping

勝又勇貴 *1 Yuki Katsumata　谷口彰 *1 Akira Taniguchi　萩原良信 *1 Yoshinobu Hagiwara　谷口忠大 *1 Tadahiro Taniguchi

*1立命館大学
Ritsumeikan university

This paper propose a new statistical semantic mapping method called *SpCoMapping*, which integrates probabilistic spatial concept acquisition and simultaneous localization and mapping via Markov random field to yield semantic information. Using a simulation environment, we compared in an experiment SpCoMapping with other previously proposed methods. We show that our method can generate better semantic maps than CNN-based semantic mapping and spatial concept formation.

1. はじめに

ロボットは，人間が実環境で使用する場所・物体・行動・人の名前などの語彙を理解して，ユーザが求めるタスクを実行する必要がある．例として，ロボットがユーザから与えられるコマンド「キッチンを清掃して」や「ジョセフの部屋のゴミを拾って」などを実行するためには，ロボットがその環境の地図上において「キッチン」や「ジョセフの部屋」の領域を知る必要がある．空間の Semantic Mapping は，ロボットが人間とコミュニケーションをとり，要求されるタスクを適切に実行するために，環境の地図に意味情報を与えるタスクである [Kostavelis 15]．

日常生活において語彙は自宅やオフィスといった環境に依存し，記号システム自体が動的であるため，未知の語彙が含まれている [T.Taniguchi 16]．従来の Semantic Mapping に関する研究では，事前知識として場所を表す語彙のラベルを用いた研究が多く行われてきた．しかし，これらの手法では各環境に存在する未知の語彙を扱うことが困難である．様々な環境と語彙に対応し，人間との円滑なコミュニケーションを実現するために，家庭環境において動作するロボットにとって，人間の発話に含まれる未知の語彙に対応できる Semantic Mapping は重要である．

Semantic Mapping は，Simultaneous Localization and Mapping (SLAM) を用いて得られた環境の地図を場所を表す語彙のラベルを持つように拡張する研究である．Sunderhauf らは，RGB データを語彙のラベルに変換する Convolutional Neural Network (CNN) を用いた Semantic Mapping を提案した [Sunderhauf 16]．この手法では，Semantic Mapping の語彙のラベルとして CNN による画像認識の結果を使用している．しかし，この画像認識をベースとした手法には 3 つの課題がある．

1 つ目は過去のサイクルで地図に塗られたラベルを別のラベルで上書きしてしまう問題である．例えば，ロボットの位置が同じであっても，廊下から部屋に入るときと部屋を出るときに得られる画像認識の結果が異なることで，先行研究では部屋に入るときに生成されたラベルを部屋から出るときの新しいラベルで上書きしてしまうという問題が生じる．Semantic Map 上のラベルは情報を単純に上書きするのではなく，統計的に保存すべきであり，提案手法は，地図上の各セルのラベルを確率変数としてモデル化することで，この問題を解決する．

2 つ目の問題は，Semantic Map が画像情報のみに基づいている点である．人間は視覚情報のみに基づいて場所のカテゴリを形成するわけではない．環境内で形成される Semantic Map 上のカテゴリの領域や種類は，視覚情報だけでなく，位置情報，ユーザーの発話などの他の情報によっても影響されることが望ましい．提案手法では，ユーザの発話などのマルチモーダル情報を扱う確率的生成モデルの一部としてマルチモーダルカテゴリ分類を用いることでこの問題を解決する．

3 つ目の問題は，既存の Semantic Mapping では，各環境の未知の語彙を扱うことができない点である．実環境において考えられる全ての場所を表す語彙とその特徴量を訓練データとしてロボットに与えることは困難である．これに対し提案手法では，未知語を扱う階層的ベイズモデルに基づく教師なし学習でこの問題を解決する．

ロボットが場所の語彙を獲得するのと同時に，カテゴリと場所の領域を推定する手法として場所概念獲得に関する研究がある [A.Taniguchi 16,Ishibushi 15]．Taniguchi らは，ロボットが未知語を段階的に獲得することを可能にする non-parametric Bayesian spatial concept acquisition method (SpCoA) を提案した [A.Taniguchi 16]．これらの研究では，マルチモーダル情報を扱うことができ，また未知の語彙を獲得することが可能である．しかし既存の場所概念獲得に関する研究では，位置分布がガウス分布によってモデル化されているため，環境の地図への適切な Semantic Mapping を行うことができない．そのため，形成された Semantic Map 上の領域において，環境の様々な形状，例えば壁や障害物による領域の分断などが考慮されていないという問題がある．

本稿では，マルコフ確率場 (Markov Random Field: MRF) と場所概念獲得の手法 [Ishibushi 15, A.Taniguchi 16] を統合した Semantic Mapping 手法である SpCoMapping を提案する．SpCoMapping には，上記の問題を解決する以下の特徴がある．

1. Semantic Map の各セルが確率変数を持ち，ラベルを上書きせず確率的に求めることができる．
2. MRF によって，Semantic Map 上の各領域は環境の形状を考慮できる．
3. 視覚情報だけでなく，人間とロボットとの対話によって得られる語彙などの情報，つまりマルチモーダル情報に基づいて Semantic Map が形成される．

連絡先: 勝又勇貴，立命館大学情報理工学研究科，滋賀県草津市野路東 1-1-1，yuki.katsumata@em.ci.ritsumei.ac.jp

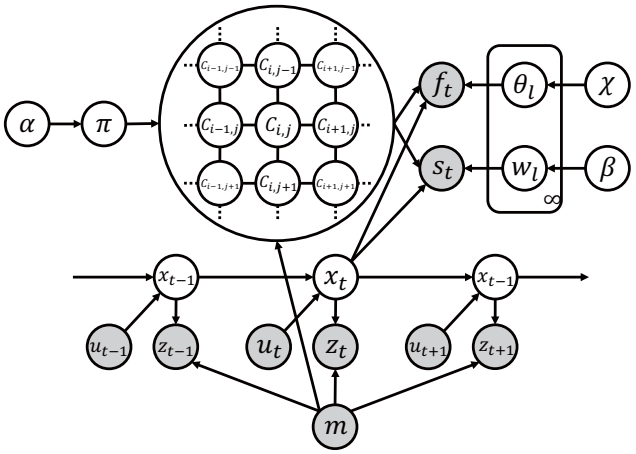


図 1: SpCoMapping のグラフィカルモデル

表 1: SpCoMapping のグラフィカルモデルの変数

m	環境の地図
x_t	ロボットの自己位置
u_t	制御情報
z_t	深度情報
$C_{i,j}$	セル (i,j) の場所概念のインデックス
f_t	画像特徴量
s_t	語彙の特徴量 (bag-of-words)
π	$C_{i,j}$ の多項分布のパラメータ
θ_l	f_t の多項分布のパラメータ
w_l	s_t の多項分布のパラメータ
α, β, χ	事前分布のハイパーパラメータ

4. 音声認識器の辞書に含まれる語彙であれば、新たな場所の名前として学習することができる．
5. Dirichlet process (DP) [Teh 05] をカテゴリ数の事前分布として使用し，場所概念のカテゴリ数を推定できる．

本稿では，SpCoMapping[*1] によって，上記の問題を解決できることをシミュレーション環境で示す．

2. 提案手法

2.1 概要

SpCoMapping のグラフィカルモデルを図 1，各変数の説明を表 1 に示す．まず，ロボットは自己位置推定を用いて環境を移動し，RGB データを取得する．ロボットが移動する間，人間は場所を表す語彙をロボットに与える．SpCoMapping では，Sunderhauf ら [Sunderhauf 16] と同様に事前に訓練された CNN を用い，画像特徴量を得る．また，人間が発話した音声を音声認識システムで認識し，その認識結果の Bag-of-Words (BoW) を語彙の特徴量として用いる．次に，ロボットはマルチモーダル情報を統合することによって場所概念を学習し，確率的生成モデルを用いて Semantic Map を形成する．

2.2 MRF

本手法では従来の場所概念学習においてガウス分布で表現されていた位置分布を，MRF によって表現する．環境の地図の各セルを MRF のノードとして用いることで，ロボットは環境の地図の形状（壁や障害物など）を考慮できるようになる．マルコフ確率場は以下の式で表される．

$$p(C_{i,j}|C_{\partial i,j};\gamma) \quad \propto \quad \exp\left\{\sum_{r\in\partial i,j}\gamma\,\delta(C_{i,j},C_r)\right\} \qquad (1)$$

ここで，$C_{\partial i,j}$ は $C_{i,j}$ の隣接するノード，γ は相互作用係数，$\delta()$ はデルタ関数を示す．

3. 実験

SpCoMapping の Semantic Mapping の性能を評価し，既存の方法と比較する実験を行った．定量評価を行うため，SIGVerse[*2] を用い，複数の家庭環境を模したシミュレーション環境において実験を行った．本実験において SpCoMapping に使用したデータセットは Github に公開した [*3]．

3.1 実験条件

画像特徴量を得るための CNN の実装には，CNN のフレームワークである Caffe [Jia 14] を使用した．また，CNN の訓練データとして，Places205 dataset [Zhou 14] を使用した．CNN は AlexNet [Krizhevsky 12] を使用した．語彙情報をロボットに与える際，Semantic Mapping を評価することに焦点を当てるため，音声認識を使用せずに，ロボットにテキストデータを直接与えた．実験で比較したのは以下の 5 つの方法である．

(A) カテゴリ数が未知の場合の SpCoMapping (DP)

(B) カテゴリ数が既知の場合の SpCoMapping (Dir)

(C) Spatial concept formation に語彙の学習を追加した手法 [Ishibushi 15]

(D) Sunderhauf らの CNN-based semantic mapping [Sunderhauf 16]

(E) Nearest neighbor

手法 (A) では，DP に week-limit 近似 [Fox 11] を用い，場所概念の上限数を 120 と設定した．手法 (B) では，場所概念の数を正解ラベルのカテゴリ数と同じ値に設定し，ディリクレ分布から場所概念の多項分布のパラメータのサンプリングを行った．手法 (E) は最近傍法である．この手法は位置的に最も近い語彙のラベルと同じラベルを獲得する手法である．Ishibushi らの手法 [Ishibushi 15] は，画像特徴量と自己位置情報のみを使用しているが，実験では語彙情報を扱うモデルを手法 (C) として使用した．手法 (D) は，占有格子地図上のフリースペースのセル全てに対してはカテゴリ分類を行わない．そこで，全てのピクセルにラベルを与えるために手法 (D) を用いて Semantic Map を形成した．図 2 は，各手法によって生成された Semantic Map を示している．

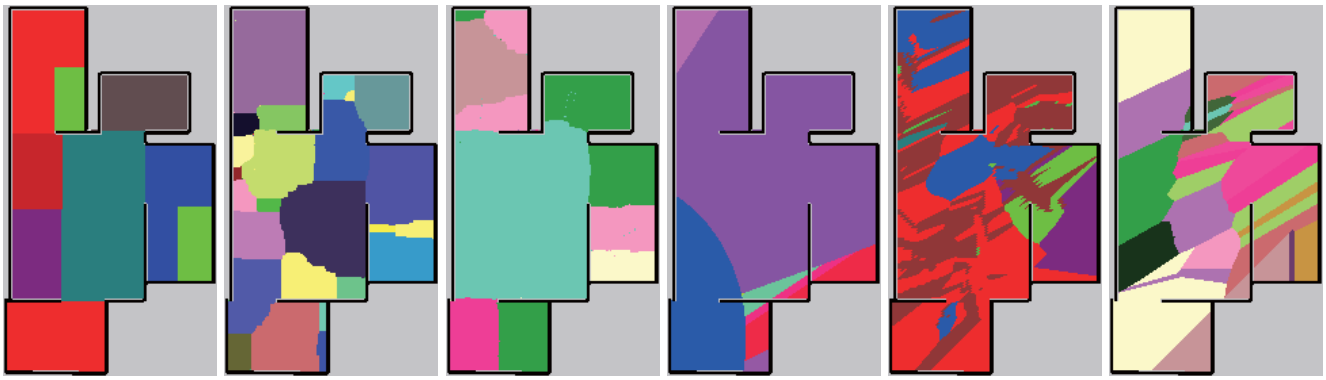
正解ラベルは，任意に選ばれた参加者が 3D シミュレータからの情報と 2D マップを参照して，各環境の Semantic Map を描画することで作成した．実験では，各環境で使用されると考えられる場所を表す語彙をロボットに与えた．用いた語彙のリストを，表 2 に示す．語彙のリスト内で下線を引いた語彙は，CNN のラベルには存在しないものを表している．また，これらの語彙には正解ラベルで用いられていない語彙も含まれる．

SpCoMapping は場所概念のカテゴリ数が未知の場合と既知の場合，いずれも 5000 イテレーション実行する．カテゴリ数が

∗1 Source code on Github: https://github.com/EmergentSystemLabStudent/spco_mapping.git

∗2 SIGVerse: http://www.sigverse.org/wiki/en/

∗3 Dataset on Github:https://github.com/EmergentSystemLabStudent/Dataset_of_SpCoMapping.git



(a) 正解ラベル (b) Proposed method (DP) (c) Proposed method (Dir) (d) The spatial concept [Ishibushi 15] (e) CNN-based semantic mapping [Sunderhauf 16] (f) Nearest neighbor

図 2: Room2ldk4 において形成された Semantic Map の例: (a) は人間が生成した正解ラベル，(b) は SpCoMapping の場所概念のカテゴリ数が未知の場合，(c) は SpCoMapping の場所概念のカテゴリ数が既知の場合，(d) は既存の場所概念獲得の Semantic Mapping 結果，(e) は CNN ベースの Semantic Mapping の結果，(f) は Nearest Neighbor の結果である．それぞれの Semantic Map の色の違いはそれぞれの Map におけるカテゴリの違いを示している．

表 2: シミュレーション環境で使用した語彙のリスト

bedroom	shower	kitchen	entrance	living
dining	closet	corridor	parlor	
window	tree	shelf	chair	sofa
washroom	toilet	refrigerator	TV	wall
bed	table	bath	door	

未知の場合のハイパーパラメータは $\alpha = 1.0 \times 10^8$，$\beta = 500$，$\chi = 1.0 \times 10^4$，$\gamma = 4.0$ とする．カテゴリ数が既知の場合のハイパーパラメータは $\alpha = 5.0 \times 10^7$，$\beta = 100$，$\chi = 1.0 \times 10^4$，$\gamma = 3.0$ とする．使用したハードウェアは，Ubuntu 14.04 LTS 64-bit，ROS Indigo with 31.4 GB memory，Intel Core i7-4770K CPU @ 3.50 GHz × 8，Gallium 0.4 on NVE4 である．

3.2 結果

3.2.1 クラスタリング

各手法に対して Adjusted Rand Index (ARI) [Hubert 85] を計算した．ARI は，2 つのクラスタ間の類似性の尺度である．2 つのクラスタが同じ場合は ARI は 1 となり，ランダムである場合は ARI は 0 となる．Semantic Mapping は，地図上のピクセルをクラスタリングするタスクと見なすことができるため，ARI を用いて提案手法と既存手法の性能を比較した．

結果を表 3 に示す．SpCoMapping は，他の手法と比較して各環境で高い値を示した．このことから，SpCoMapping は 1 章で述べた上書きの問題の解決と環境に合わせた任意の形状を持つ領域を生成することができることを示している．つまり，SpCoMapping によって生成された Semantic Map のカテゴリは，人によって生成された Semantic Map のカテゴリに近いと言える．

3.2.2 単語から場所の予測

ナビゲーション，部屋の掃除，探しものなどのユーザとのコミュニケーションが必要なタスクをロボットが行う場合，ロボットは場所を表す語彙からユーザの指示する場所を推定する必要がある．そこで，各手法で推定した場所の正解率を比較した．正解率の計算は次のように計算する．

$$R_{match} = \sum_{s \in V}^{|V|} \frac{\sum_{i=1}^{M} \delta(s_i, L_i)}{M} \tag{2}$$

ここで V は正解ラベルに含まれる語彙の数，M は占有格子地図中のフリースペースの数，s_i は Semantic Map のピクセル i に与えられた語彙，L_i は正解ラベルのピクセル i に与えられた語彙を表す．

SpCoMapping において，語彙 s を与えられたときの場所概念のインデックスの推定は以下の式で行う．

$$C_s = \underset{C}{\operatorname{argmax}}\, p(C \mid s_t = s, \pi, w) \tag{3}$$

ここで s_i は Semantic Map 中のピクセル i における $C_i = C_s$ のときの語彙を表す．

結果は表 4 に示す．手法 (B) は，他の方法よりも高い値を示した．手法 (A) の値は，Room1dk5 と Room1dk4 において 5 つの手法の中で最も良い結果が得られた．しかし，ピクセル数の大きな環境（Room1dk6 と Room2ldk4）の結果は他の手法に劣っている．この 2 つの環境で SpCoMapping の性能が低下した原因はカテゴリ数の推定が間違っていたためである．カテゴリ分類のノンパラメトリックベイズ推定は，不安定であるため，SpCoMapping (Dir) は SpCoMapping (DP) よりも安定した結果が得られることがわかる．

4. 結論

本稿では，MRF を用いて場所概念獲得手法を拡張し，SpCoMapping と呼ばれる Semantic Mapping について述べた．シミュレーション環境における実験により，SpCoMapping は既存の Semantic Mapping の問題に対処できることが示された．さらに実験において，SpCoMapping によって生成された Semantic Map は，人間によって生成された Semantic Map と，既存の手法よりも一致している結果が得られた．また，SpCoMapping により生成された Semantic Map は，単語入力からの場所推定，すなわち人間とロボットとのコミュニケーションの観点から既存の方法よりも優れていることを示した．

表 3: ARI の結果

Method	Room1dk5	Room1dk6	Room1ldk4	Room1ldk5	Room2ldk4
(A) SpCoMapping (DP)	**0.5356**	0.4548	**0.5877**	**0.4925**	0.4623
(B) SpCoMapping (Dir)	0.4471	**0.5355**	0.4393	0.3354	**0.4963**
(C) Spatial concept formation [Ishibushi 15]	0.3524	0.4692	0.2657	0.3740	0.2883
(D) CNN-based semantic mapping [Sunderhauf 16]	0.3102	0.2742	0.2505	0.3067	0.2559
(E) Nearest neighbor	0.3371	0.2870	0.4830	0.3337	0.3968

表 4: 領域の正解率

Method	Room1dk5	Room1dk6	Room1ldk4	Room1ldk5	Room2ldk4
(A) SpCoMapping (DP)	**0.2612**	0.0268	**0.2185**	0.1338	0.0001
(B) SpCoMapping (Dir)	0.2071	**0.6368**	0.1461	0.1119	0.1780
(C) Spatial concept formation [Ishibushi 15]	0.0463	0.2587	0.2154	**0.3200**	**0.1876**
(D) CNN-based semantic mapping [Sunderhauf 16]	0.1174	0.0761	0.0949	0.0766	0.1245
(E) Nearest neighbor	0.1260	0.1607	0.1729	0.1252	0.1292

今後の研究では，SpCoMapping を人間とのコミュニケーションを必要とするタスク，例えば「私の部屋をきれいにしてください」と言ったようなタスクなどに適用することを考えている．また，シミュレーション実験で明らかになった SpCoMapping (DP) の安定性を向上させることも課題の一部である．さらに，SpCoMapping はバッチ学習を採用しているため，SpCoMapping をオンライン学習が行えるように改善し，環境の地図が未知の環境において動作するために SLAM と統合することも今後の課題として挙げられる．

いくつかの課題は残っているが，提案手法は教師なし学習に基づく空間の Semantic Mapping の性能を大幅に改善し，実環境における人間の発話を Semantic Mapping に利用できることを示した．本研究は，近い将来，家庭環境において能動的に学習するロボットと人のコミュニケーションを円滑化することに貢献すると考えられる．

参考文献

[Kostavelis 15] Ioannis Kostavelis, Antonios Gasteratos, "Semantic mapping for mobile robotics tasks: A survey", Robotics and Autonomous Systems, 2015, pp.86-103

[T.Taniguchi 16] Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh, "Symbol Emergence in Robotics: A Survey, Advanced Robotics", Advanced Robotics vol.30 no.11-12, 2016, pp.706-728

[Sunderhauf 16] Niko Sunderhauf, Feras Dayoub, Sean McMahon, Ben Talbot, Ruth Schulz, Peter Corke, Gordon Wyeth, Ben Upcroft, and Michael Milford, "Place Categorization and Semantic Mapping on a Mobile Robot", International Conference on Robotics and Automation (ICRA), 2016, pp.5729-5736

[A.Taniguchi 16] 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", Transactions on Cognitive and Developmental System, 2016, pp.285-297

[Ishibushi 15] Satoshi Ishibushi, Akira Taniguchi, Toshiaki Takano, Yoshinobu Hagiwara and Tadahiro Taniguchi, "Statistical Localization Exploiting Convolutional Neural Network for an Autonomous Vehicle", IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, 2015, pp.1369-1375

[Teh 05] Yee Whye Teh, Michael I. Jordan, Matthew J. Beal and David M. Blei, "Sharing clusters among related groups: Hierarchical dirichlet processes", Advances in neural information processing systems (NIPS), 2005, pp.1385-1392

[Jia 14] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding", Proceedings of the 22Nd ACM International Conference on Multimedia, 2014, pp.675-678

[Zhou 14] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva, "Learning Deep Features for Scene Recognition using Places Database", Advances in Neural Information Processing Systems 27, 2014, pp.487-495

[Krizhevsky 12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems 25, 2012, pp.1097-1105

[Fox 11] Emily B Fox, Eric B Sudderth, Michael I Jordan and Alan S Willsky, "A sticky HDP-HMM with application to speaker diarization", Institute of Mathematical Statistics, 2011, pp.1020-1056

[Hubert 85] Lawrence Hubert, "Comparing Partitions", Journal of Classification Volume 2, Issue 1, 1985, pp.193-218