

# 確率的生成モデルを用いたロボットの空間概念と音声言語の獲得

Acquisition of Spatial Concepts and Spoken Language using Probabilistic Generative Model by Robot

2167005



研究代表者

立命館大学  
総合科学技術研究機構

日本学術振興会  
特別研究員(PD)

谷口 彰

## [研究の目的]

人間と共存し様々な環境で動作するロボットは、周囲の環境を認識し適応的に概念や言語を学習・活用することが求められる。ロボットの概念・言語獲得の研究は、単語知識や構文知識を手で組み込んだ自然言語処理、大語彙辞書を利用した音声認識などとは違い、事前に知識を持たない中で音声インタラクションを通して概念や言語をボトムアップに獲得していかなければならない。ロボットが人および環境との相互作用によって概念や言語を獲得するためには、センサのノイズ、移動の誤差、音声認識誤りなどの多くの不確実性への対処が重要である。

本稿では、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する。また、ロボットが場所や空間に関するマルチモーダルなセンサ情報に基づき自律的に形成したカテゴリ知識のことを場所概念と定義する。

本研究は、ロボットが人の連続音声発話から自律的に場所概念を獲得し、より効果的な空間認知と自己位置推定を可能にすることを目的とする。さらに、教師なし学習に基づく場所概念獲得に関する手法を応用し、ロボットと人間との空間動作に関わる言語コミュニケーションの実現を目指す。本研究では主として、①音節認識誤りあり発話文からの高精度な未知語の単

語分割、②場所概念と言語モデルの相互学習の方法、③画像や物体の情報を統合した場所概念の獲得について問題とする。

## [研究の内容、成果]

### 1. 発話文からの高精度な語彙獲得

これまでの研究成果として、音節認識誤りを含んだ言語情報と不確実なセンサ情報に基づく位置推定情報のベイズ確率的生成モデルおよび、ノンパラメトリックベイズ場所概念獲得モデル(SpCoA)を提案している[1]。しかし、単語分割に場所の情報が考慮されていなかったために、場所の名前が途中で分割されてしまう場合があるという問題があった。そのため、本研究では、図1に示すように場所概念と言語モデルの相互推定を行うことで、発話文からの高精度な語彙獲得を実現するSpCoA++を提案した。具体的には、SpCoA++では発話文の教師な

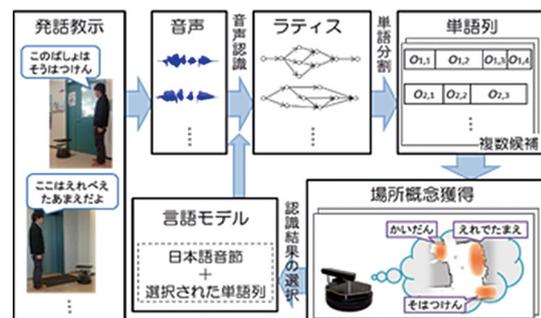


図1 場所概念と言語モデルの相互推定

し単語分割の結果に対し学習された場所概念を考慮することを行う。提案手法は SpCoA によって場所概念を獲得し、獲得した場所概念を使用して言語モデルを更新することを繰り返す。そして、特定の場所と関連性の高い単語の含まれた単語分割結果により言語モデルを更新することで語彙獲得の改善を図る。また、場所概念と単語の間の相互情報量を計算することにより、特定の場所に関連付いた特定の単語を判別することを可能とした。相互推定において使用する言語モデルは単語辞書の形式で表されるものとし、初期の言語モデルは日本語音節のみを持つものとする。

#### 〈相互推定の手順〉

以下に、SpCoA++ の場所概念と言語モデルの相互推定の手順を述べる。また、相互推定の流れを表す図を図 1 に示す。

##### (1) 言語モデルを用いた発話文の音声認識

ロボットは現在の言語モデルを用いて連続音声発話の音声認識を行う。音声認識結果はラティス (WFST のグラフ) 形式で得る。

##### (2) 音声認識ラティスを教師なし単語分割

音声認識ラティスの教師なし形態素解析手法である latticelm によって音声認識結果のラティスを用いて教師なし単語分割を行う。教師なし単語分割のパラメータ候補を複数用意し、様々な単語分割結果を得る。様々な単語分割の候補を考えることにより、より場所と関連性の高い単語の切り出し方をした単語分割結果を選びとることが可能になる。

##### (3) 場所概念の学習

単語分割の候補ごとに、SpCoA を用いて場所概念の学習を行う。

##### (4) 単語分割結果の選択

単語分割結果の選択は、 $t$  時刻ごとの単語  $S_{t,b}$  と場所概念の index  $C_t$  についての相互情報量を計算し、最も高い値のものを選ぶ。また、提案手法では、より高精度な語彙獲得のために、二値化した変数に基づく相互情報量を利用し場所に関係した単語の選択を行う。

##### (5) 言語モデルの更新

言語モデルの更新は、日本語音節のみを登録した音声認識器の単語辞書に選択された単語分割結果の全単語を追加する形で行う。

##### (6) 繰り返し推定

(1)~(5) を複数回繰り返す。

##### (7) 繰り返し終了後

最後に選ばれた単語分割結果による学習結果を最終的な場所概念の学習結果とする。同様に、言語モデルも最後の更新結果を最終的に推定された言語モデルとする。

さらに、提案手法を実機ロボットに実装し、語彙獲得の高精度化実現の検証のために、実環境上で場所概念獲得の実験を行い、相互推定手法の有効性を示した。実験結果の一例を図 2 に示す。この成果は主に査読付き国際誌 [5] にて発表済みである。

## 2. 画像や物体情報を用いた場所概念獲得

これまでの提案モデル (SpCoA, SpCoA++) では、場所概念は位置情報と言語情報により獲得していた。人間の空間認知を考えた場合、視覚から得られる風景や物体の配置といった情報は空間や場所を認識する上で極めて重要な要素である。視覚情報を使わない場合では、ロボットは「どこ」に「なに」があるのかを未知環境から学習することはできない。そのため本研究では、場所概念獲得モデルを拡張する形で画像や物体概念などと統合し、マルチモーダルな情報を活用することを行った。マルチモーダルな情報を統合することにより、位置情報のみでは認識不可能だった風景の違いや

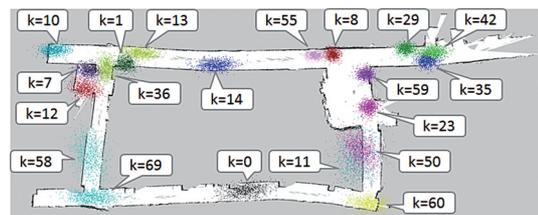


図 2 場所概念の学習結果

物体の位置などの場所ごとの特性の学習，判別が可能になる。

本研究では，視覚から得られる風景や物体の配置といった情報を場所概念獲得に利用するために，場所概念獲得のモデルに視覚情報を追加することを行った。まず，[2] では，位置と言語と画像というマルチモーダルな情報に基づく場所概念獲得のモデルを提案した。これにより，ロボットが場所の名前，空間領域，風景情報の相互補完的な推定を可能とした（図3）。次に，[3] では，Faster RCNN や YOLO によって得られる物体認識情報によるカテゴリ形成の影響度を調査した。また，[4] では，場所概念と物体概念を統合した概念形成モデルの構築のための基礎となる成果を得た。

本稿では，特に [2] の研究成果について紹介する。[2] では，Simultaneous Localization And Mapping (SLAM) と位置・音声言語・画像のマルチモーダル情報に基づくノンパラメトリックバイズ場所概念獲得モデルを統合した SpCoSLAM を提案している（図4）。また，環境の地図・場所概念・語彙を同時に学習できるオンライン学習アルゴリズムの構築を行った。

SLAM は，自己位置推定と地図生成を同時に行う問題である。特に FastSLAM では，Rao-Blackwellized Particle Filters (RBPF) によるオンラインアルゴリズムによって効率的な自己位置推定と地図生成を実現している。本研究では，FastSLAM アルゴリズムを場所概念獲得の確率的生成モデルに対し適用することを

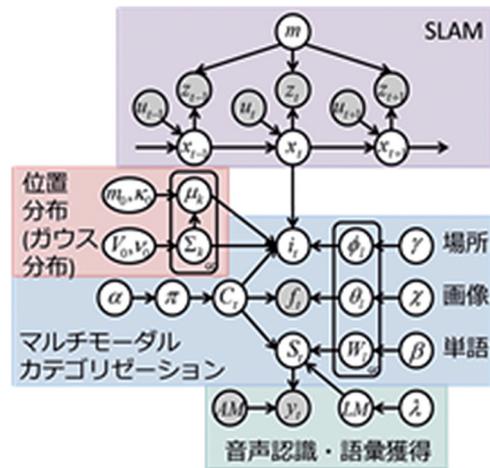


図4 SpCoSLAM のグラフィカルモデル

行う。SpCoA は Particle Filter に基づいた自己位置推定手法である MCL に語彙獲得を統合したモデルであるため，SLAM への拡張は自然に行うことができる。これにより，ロボットは環境の地図を生成すると同時に場所に関する語彙を逐次的に獲得することが可能となる。

実験では，地図を持たない未知環境下からでも場所概念の獲得が可能であること，環境中の未探索の領域に対して追加学習が可能であることを示した。また，比較実験により，画像情報を用いないこれまでのモデルよりも画像情報を活用した本提案モデルの方がより正確に場所を分類できることが示された。

[今後の研究の方向，課題]

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持つことなく獲得することが可能である。例えば，オフィスで客人が「どこどこ」へ連れて行ってください」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」などと返答を返すことや，場所についての知識と物体

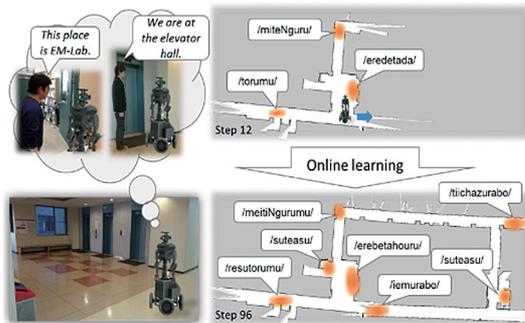


図3 ロボットによる地図・場所概念・語彙の逐次学習のイメージ図

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。

場所概念のオンライン学習の利点の一つとして、環境や場所の名前の変化に柔軟に対応できることが挙げられる。また、ロボットが誤って学習した場所概念の逐次的な修正も可能であると考えられる。したがって、人とロボットとの空間に関わる長期的な言語コミュニケーションの実現に大きく貢献することが期待される。

しかしながら、このアルゴリズムでは過去のすべてのデータを保持しておかなければならず、学習が長期に及ぶと計算コストが増大するという問題点があった。また、このアルゴリズムによる場所の推定精度はバッチ学習の6割ほどと不十分であった。これに関しては、活性化や忘却などを取り入れることにより、オンライン学習のアルゴリズムにおいて推定精度の改善と計算量の削減を図り、SpCoSLAMの頑強かつ軽量のオンラインアルゴリズムの開発を行う予定である。将来的には、動的な環境における場所概念の長期的な学習や場所概念を新規な環境への逐次的な転移学習も行っていきたい。

[成果の発表, 論文等]

1. 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, *IEEE Transactions on Cognitive and Developmental Systems*, Vol. 8, No. 4, pp. 285-297, 2016.
2. Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi, and Tetsunari Inamura: Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2017)*, 2017.
3. Shota Isobe, Akira Taniguchi, Yoshinobu Hagiwara, and Tadahiro Taniguchi: Learning Relationships between Objects and Places by Multimodal Spatial Concept with Bag of Objects, *International Conference on Social Robotics (ICSR)*, pp. 115-125, 2017. Best Interactive Session (Finalist)
4. Akira Taniguchi, Tadahiro Taniguchi, and Angelo Cangelosi: Cross-Situational Learning with Bayesian Generative Models for Multimodal Category and Word Learning in Robots, *Frontiers in Neuro-robotics*, Vol. 11, No. 66, 2017.
5. Akira Taniguchi, Tadahiro Taniguchi, and Tetsunari Inamura: Unsupervised Spatial Lexical Acquisition by Updating a Language Model with Place Clues, *Robotics and Autonomous Systems*, Vol. 99, pp. 166-180, 2018.