マルチモーダルカテゴリゼーション
～階層ベイズモデルに基づくロボットによる概念・言語獲得～

電気通信大学　中村 友昭

概念・言語学習

- 人のように言語を獲得するロボットの実現
  - 人や環境とのインタラクションにより自律的に獲得
    - 概念形成
    - 語彙の獲得（単語辞書）
    - 語意の獲得（記号接地）
    - 文法の獲得

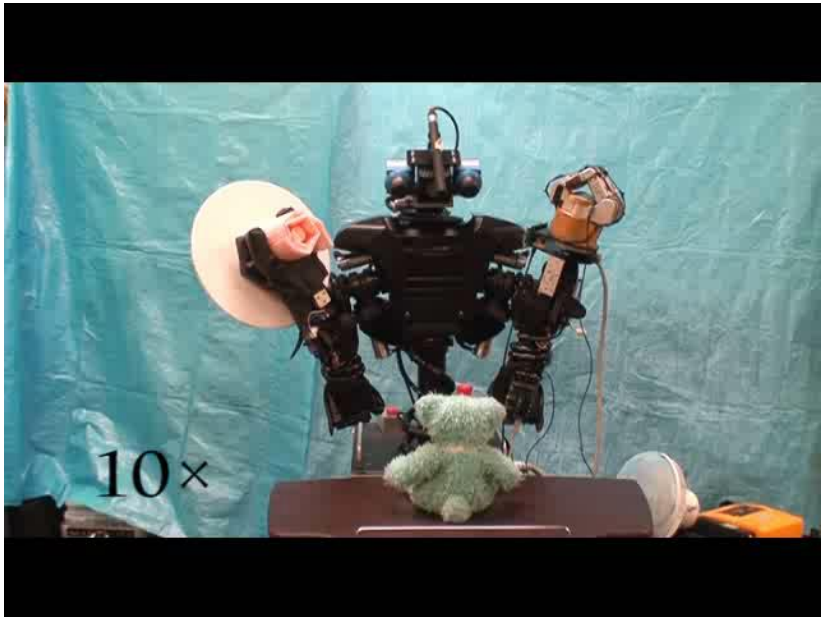
- 言語獲得アルゴリズムを確率モデルを用いて実現
  - 人間のような知能の実現
  - 人間の言語獲得過程の解明

ロボットによる概念・語意獲得

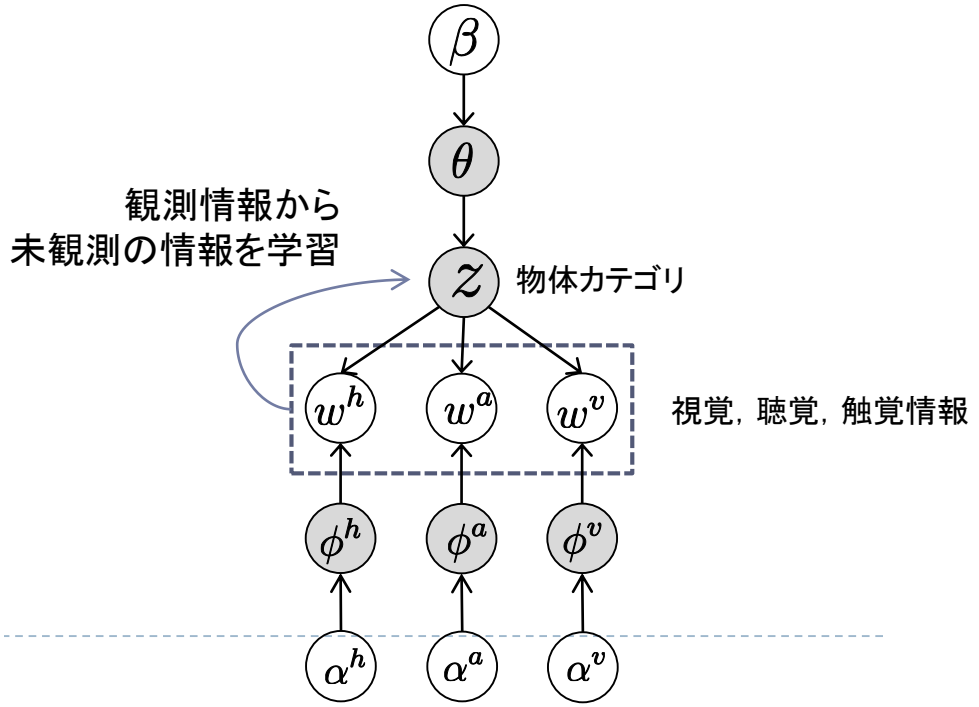
- Tomoaki Nakamura, Takaya Araki, Takayuki Nagai and Naoto Iwahashi, ″Grounding of Word Meanings in LDA-Based Multimodal Concepts″, Advanced Robotics, Vol.25, pp. 2189-2206, Apr.2012
- Takaya Araki, Tomoaki Nakamura, Takayuki Nagai, Kotaro Funakoshi, Mikio Nakano, and Naoto Iwahashi, ″Online Object Categorization Using Multimodal Information Autonomously Acquired by a Mobile Robot″, Advanced Robotics, Vol.26, pp.1995-2020, Oct.2012

概念形成

- 概念の工学的定義
  - 概念＝知覚情報のクラスタリングによって形成されたカテゴリ
  - 知覚情報：ロボットの視覚，聴覚，触覚情報
  - これらの情報を確率モデルにより教師なしで分類



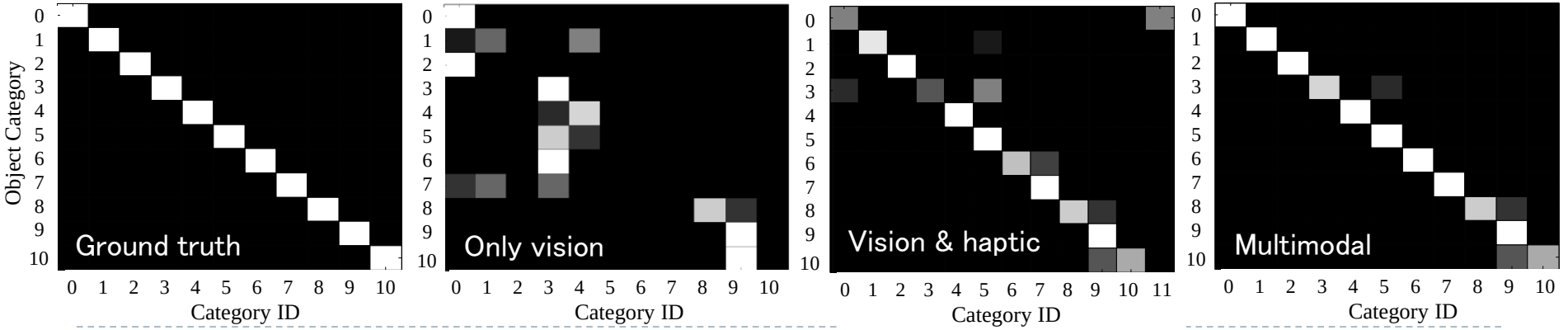
物体に見て・触れることで物体の概念を形成



概念形成

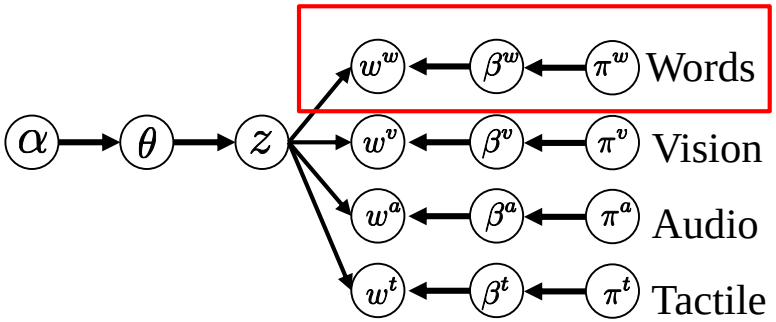
- 67個の物体から取得したマルチモーダル情報を分類

- 分類結果 （縦軸：正解カテゴリ，横軸：実際に分類されたカテゴリ，濃淡：物体の個数）

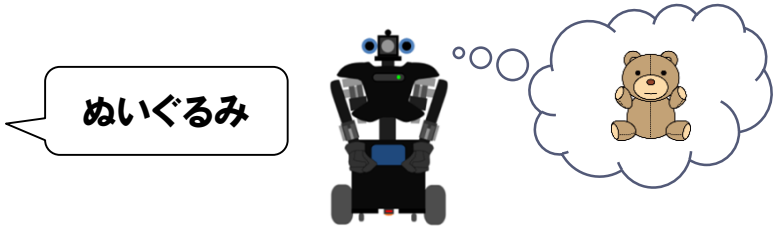


語意の獲得

- マルチモーダル情報として単語を追加
- 単語も含めたマルチモーダルな概念を形成



- 単語からマルチモーダル情報を確率的に予測可能



$$P(w^v, w^a, w^t | w^w)$$

マルチモーダル情報　　単語

単語を"感覚的"に理解可能
⇒単語の意味の理解

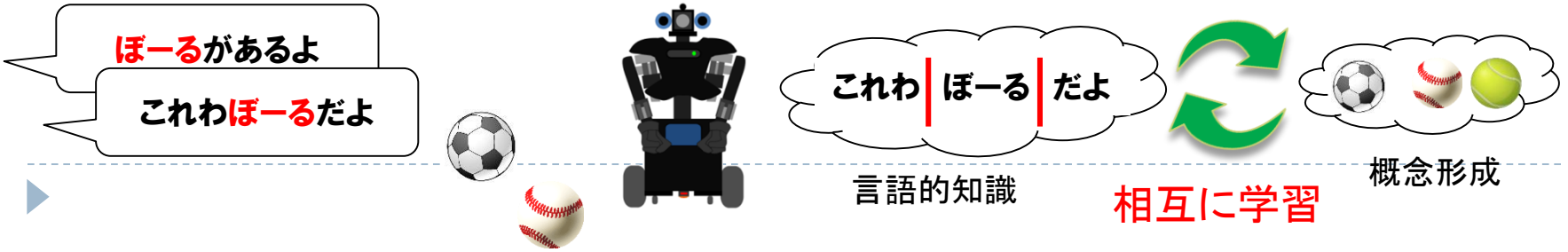
ロボットによる語彙の獲得

- Tomoaki Nakamura et al., "Mutual Learning of an Object Concept and Language Model Based on MLDA and NPYLM", IROS,2014
- Tatsuya Aoki et al., "Multimodal Learning of Object Concepts and Word Meanings by Robots", NIPS 2015 Workshop: Multimodal Machine Learning, 2015
- Joe Nishihara et al., "Online Algorithm for Robots to Learn Object Concepts and Language Model", IEEE Transactions on Cognitive and Developmental Systems, 2016

語彙の獲得

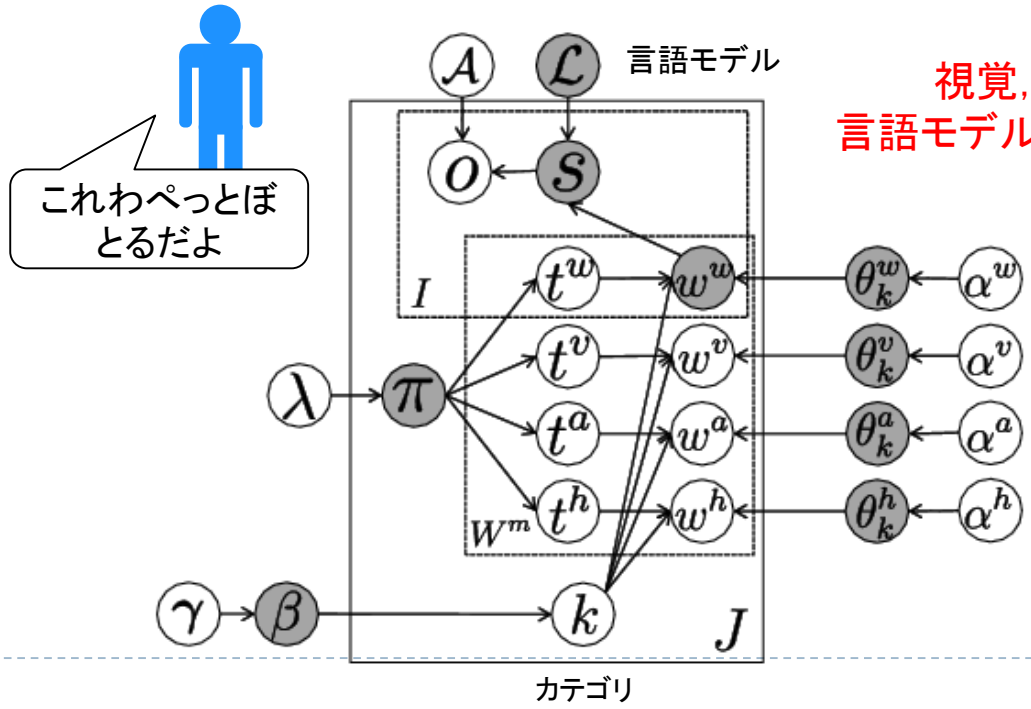
- 前の研究では語彙を持っていることが前提
  - 語彙＝音声認識・単語分割で使用する単語辞書（言語モデル）
- 語彙を獲得する際の問題
  - 語彙を持たないため
    音声が正しく認識できない
    - ぼーる？ぼーう？ごーる？
  - 単語の切れ目が分からない
    - ぼーる？わぼー？ぼーるだ？るだよ？

- これらの問題を言語的知識と概念を相互学習することで解決
  - 言語のパターンに基づく単語分割
  - 同じ概念に含まれる物体には同じ単語が教示される可能性が高い



提案モデル

- 教示音声・マルチモーダル情報の生成モデル
  - 概念の形成と同時に言語モデル（語彙）の獲得が可能
  - 物体カテゴリkと言語モデル$\mathcal{L}$が結びついたモデル➡相互に影響
    - 同じ物体には同じ単語が発話される可能性が高い
    - 同じ単語が与えられた物体は同じカテゴリの物体である可能性が高い

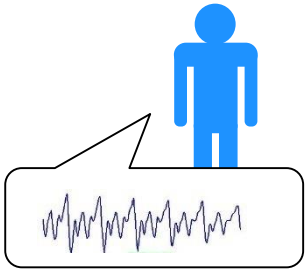


視覚，聴覚，触覚，教示音声から
言語モデルと物体カテゴリを教師なしで学習

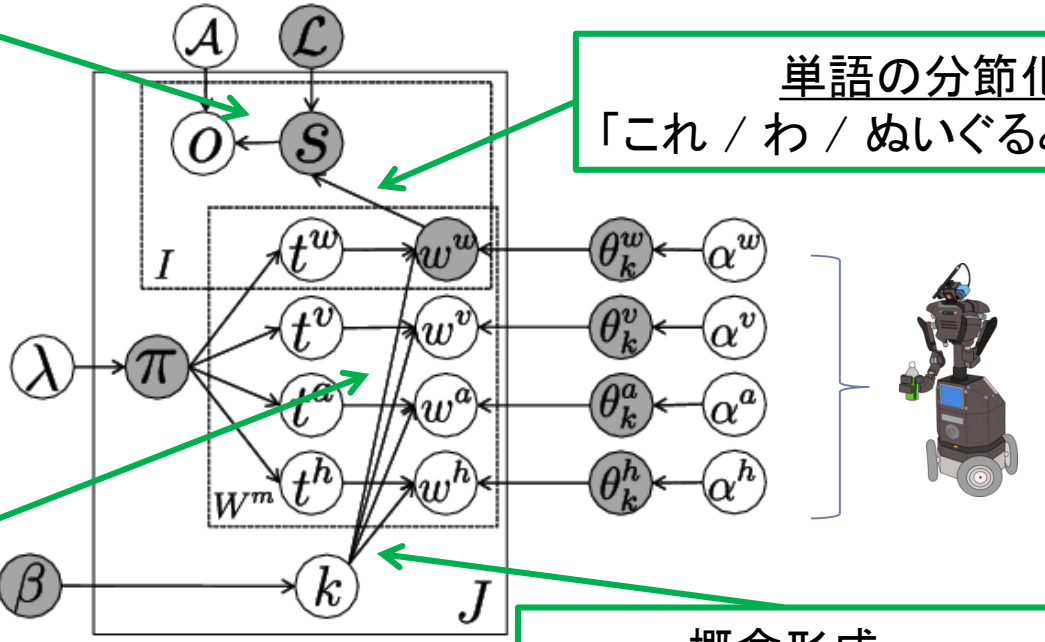
視覚，聴覚，触覚情報

モデルの構造

音声認識
「これわぬいぐるみだよ」



単語の分節化
「これ / わ / ぬいぐるみ / だよ」



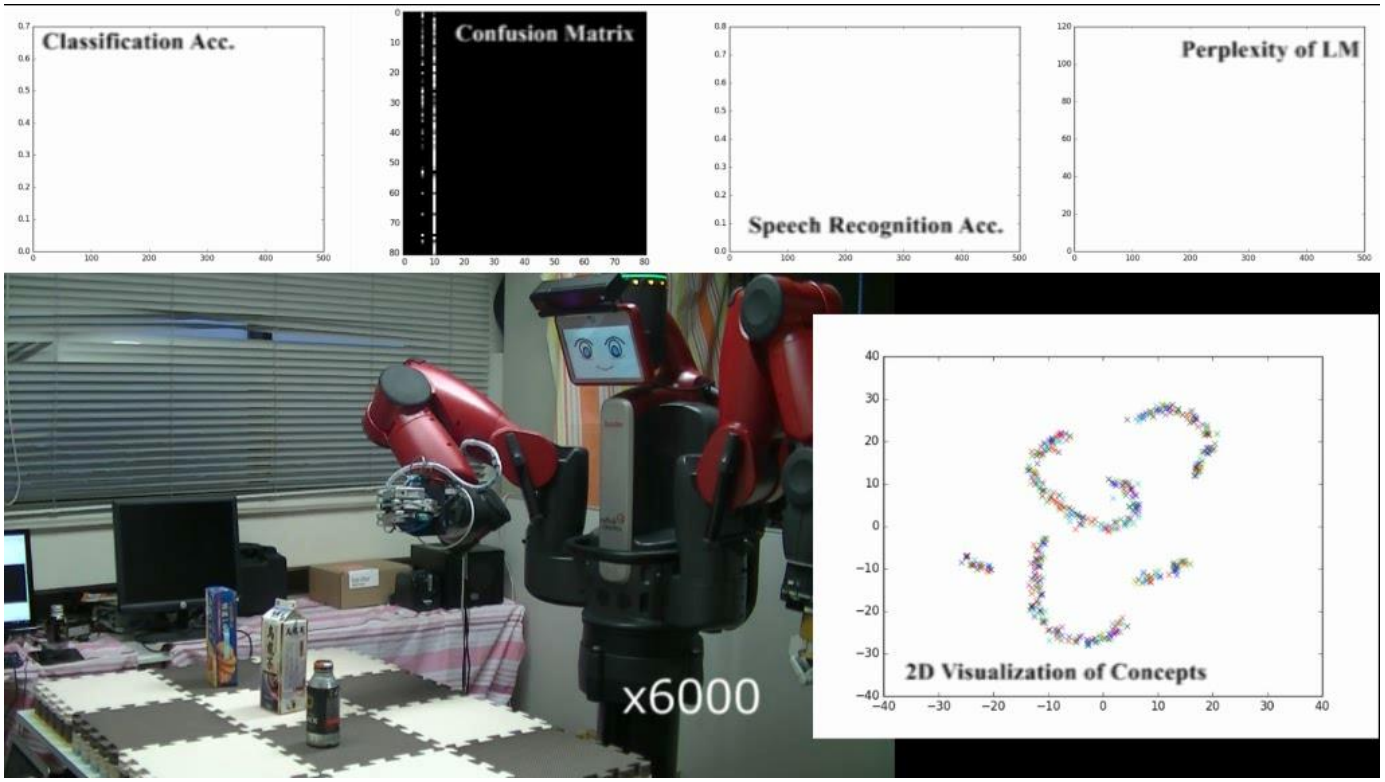
概念と単語のマッピング

概念形成

- 音声認識，単語分節化，概念形成，概念と単語の結びつきを全て**教師なし**で学習

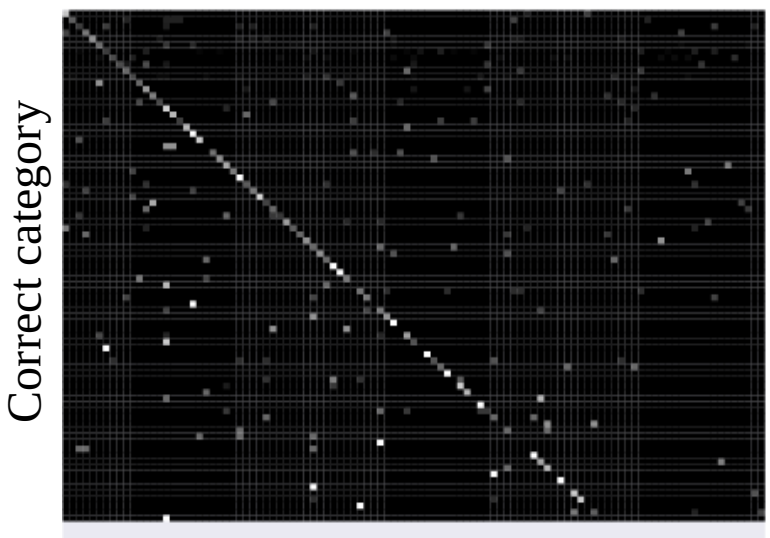
学習の様子

- 1日3～5時間程度の学習を約1ヶ月間実施(100時間強)
- 計499個の物体を学習

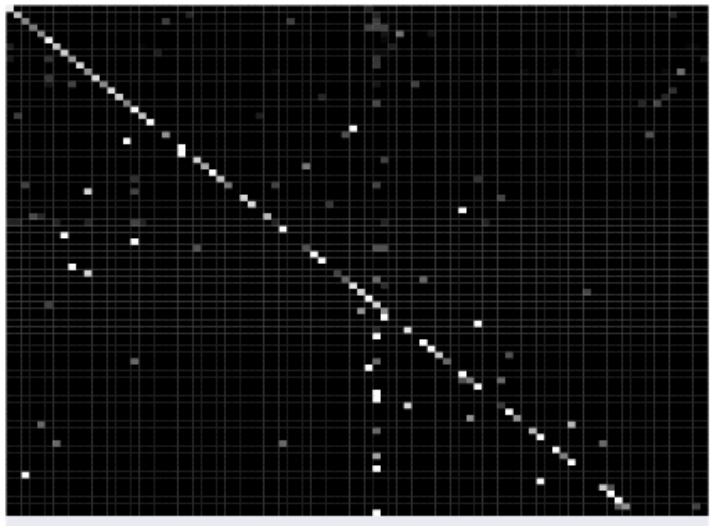


概念の形成結果

- ロボットが最終的に獲得した概念で全物体を分類



比較手法(38.1%)

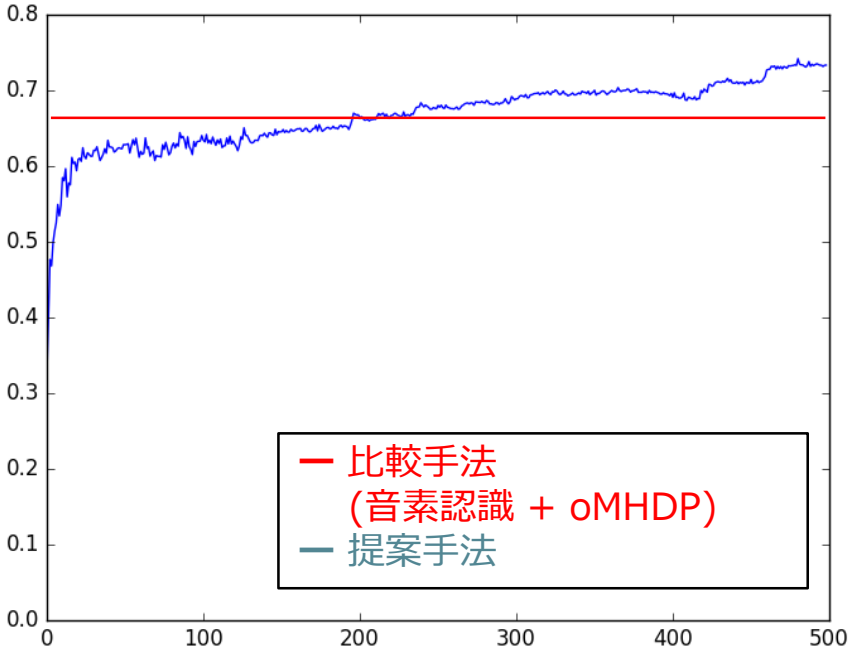


提案手法 (61.7%)

- 提案手法の方が**人に近い概念**を形成
  - 単語をより正し聞き取ることが可能になったため

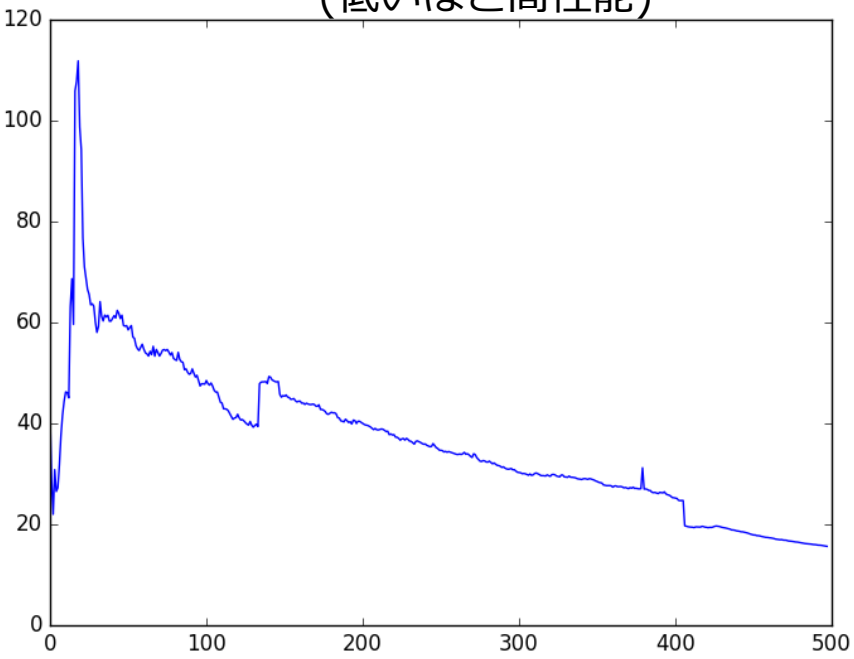
音声認識・言語モデル(語彙)の評価

音声認識精度



Number of learned objects

言語モデルのPerplexity
(低いほど高性能)



Number of learned objects

- 学習により正しい言語モデル(語彙)が獲得された
  ➡ 音声認識性能が向上

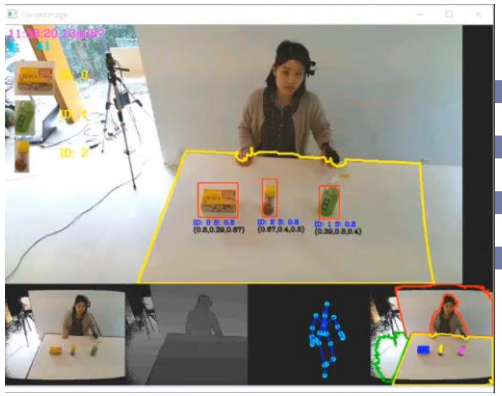
ロボットが獲得した単語

多様な概念の形成と文法の獲得

- Muhammad Fadlil, Keisuke Ikeda, Kasumi Abe, Tomoaki Nakamura, and Takayuki Nagai, "Integrated Concept of Objects and Human Motions Based on Multi-layered Multimodal LDA", IROS2013, pp.2256-2263, Nov. 2013.
- Muhammad Attamimi, Yuji Ando, Tomoaki Nakamura, Takayuki Nagai, Daichi Mochihashi, Ichiro Kobayashi and Hideki Asoh, "Learning Word Meanings and Grammar for Verbalization of Daily Life Activities Using Multilayered Multimodal Latent Dirichlet Allocation and Bayesian Hidden Markov Models", Advanced Robotics, Vol. 30, Issue 11-12, pp. 806-824, Jun. 2016

多様な概念と文法の学習

- 概念には物体だけでなく様々な概念が存在
  ➡ 人の行動シーンから物体，人，場所，動き概念を形成
- 概念クラスの順序規則を文法として学習
  ➡ 文法を学習することで観測シーンの言語表現が可能



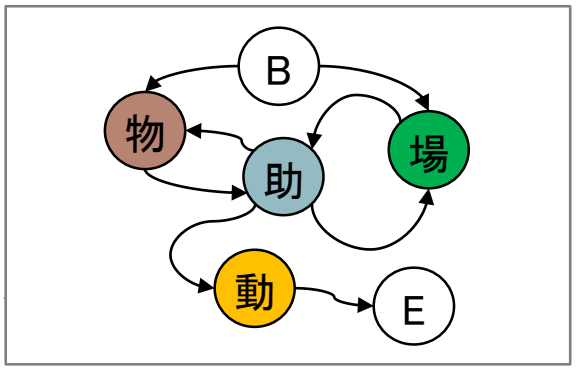
ロボットの観測シーン

物体情報
動き情報
場所情報
言語情報

概念と単語の結び付きを獲得

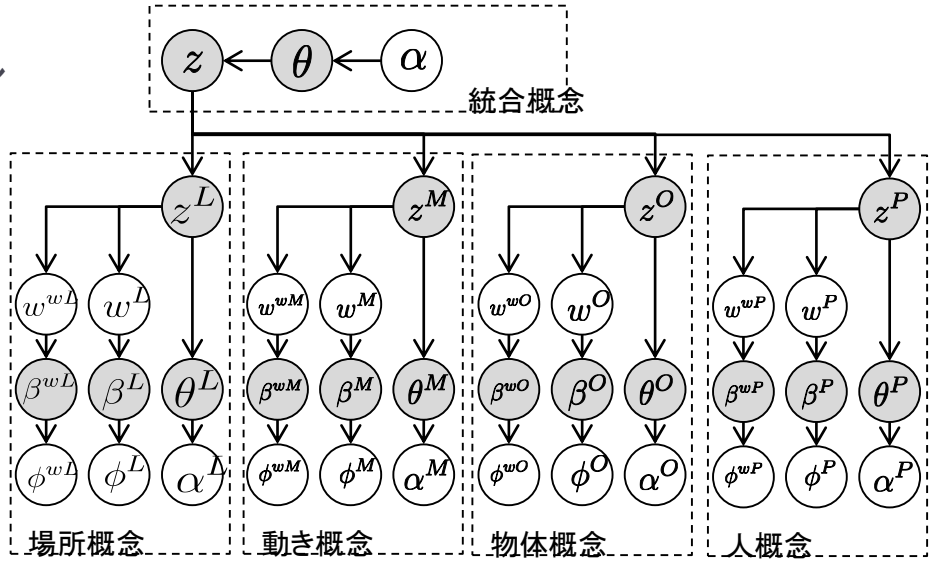
ソファ	で	ジュース	を	飲む
↓	↓	↓	↓	↓
場所	助詞	物体	助詞	動き

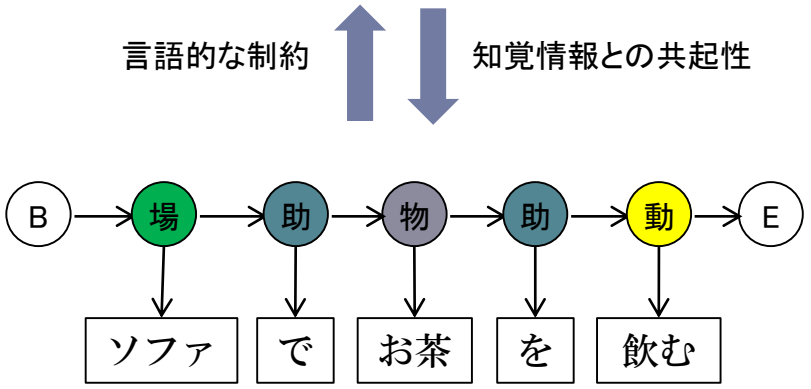
概念の遷移を文法として学習



複数概念モデルと文法モデル

- 複数概念モデル（mMLDA）
  - 場所，動き，物体，人概念の生成モデル
  - ロボットが観測した情報から概念を教師なしで学習可能
  - 単語と概念の結び付きも学習
- 文法モデル(HSMM)
  - 概念の結び付きから概念クラスの遷移順を文法として学習
  - 概念モデルの結果を初期値として教師なし学習
- 文法と概念を相互に学習
  - 言語的制約と知覚情報との共起性に基づく単語の接地





文生成

- 獲得した概念・文法からシーンを説明する文を生成
- 生成文の評価結果
  - 文法と概念の相互学習により精度が向上
- 生成例

評価基準		比較手法 (相互学習なし)	提案手法 (相互学習あり)
文法	意味		
正	正	16.06%	69.23%
正	誤	78.73%	23.76%
誤	正	0.91%	1.31%
誤	誤	4.30%	5.66%

じょせいが りびんぐで
ぽてちを たべる

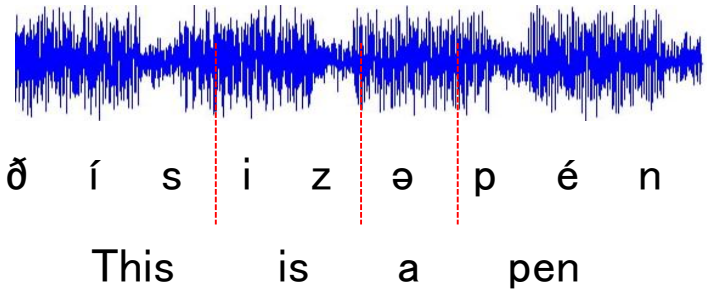
せいねんが きっちんで
そそぐ

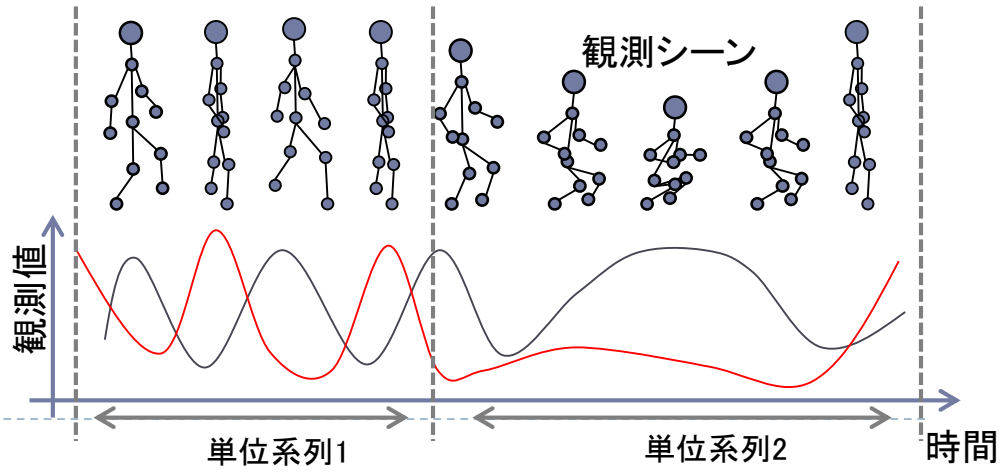
センサ情報の分節・分類に基づく概念形成

- Tomoaki Nakamura et al., "Continuous Motion Segmentation Based on Reference Point Dependent GP-HSMM", IROS2016: Workshop on Machine Learning Methods for High-Level Cognitive Capabilities in Robotics, Oct. 2016
- Tomoaki Nakamura et al., " Segmenting Continuous Motions with Hidden Semi-Markov Models and Gaussian Processes ", Frontiers in Neurorobotics (under review)

連続的なセンサ情報の分節化

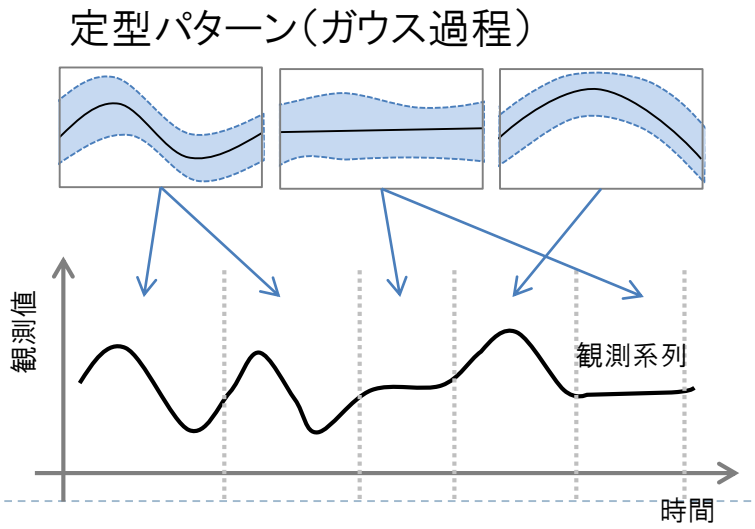
- これまでの研究では**分節化**は考えていなかった
  - 動作概念なども単位動作毎にデータを人手で区切っていた
- 実際のセンサ情報は連続であり教師なしで分節・分類する必要がある
- ロボットが連続的なセンサ情報を**教師なしで分節・分類**し概念を獲得

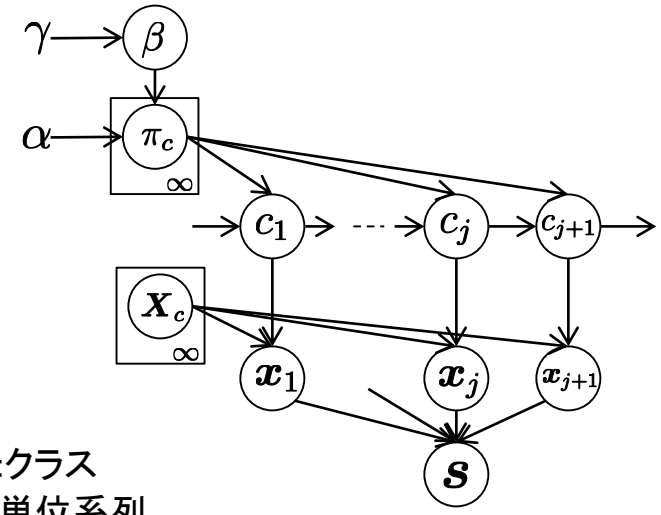




提案モデル

- 時系列データはガウス過程（GP）を出力分布とする隠れセミマルコフモデル(HSMM)によって生成されると仮定
  - ガウス過程：分節化された定型パターンを表現
  - HSMM：定型パターンの長さも隠れ変数としたHMM
- HSMMとGPのパラメータ推定することで，連続的な情報の分節・分類が可能





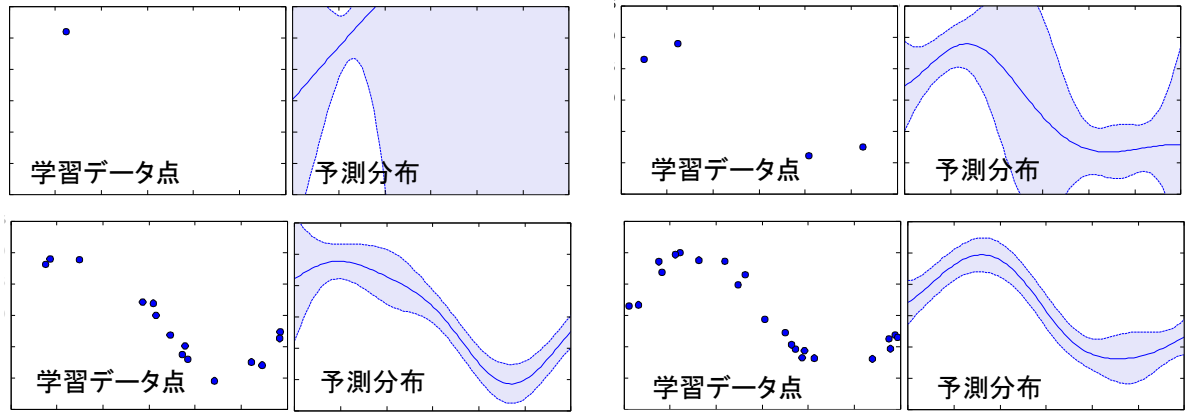
c_j :クラス
x_j :単位系列
X_c:クラスcのガウス過程のパラメータ
S :観測系列

ガウス過程

- 時刻tの出力xを予測する回帰関数の確率モデル
- 学習データ$\boldsymbol{x}$が与えられると各時刻t'での予測分布がガウス分布で得られる

$$p(x'|t', \boldsymbol{x}, \boldsymbol{t}) \propto N(x|\boldsymbol{k}^T \boldsymbol{C}^{-1} \boldsymbol{i}, c - \boldsymbol{k}^T \boldsymbol{C}^{-1} \boldsymbol{k})$$

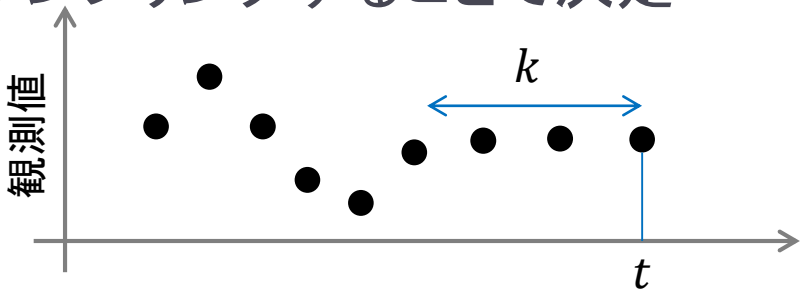
  - 軌道の一致度，観測値の曖昧性を確率で評価可能

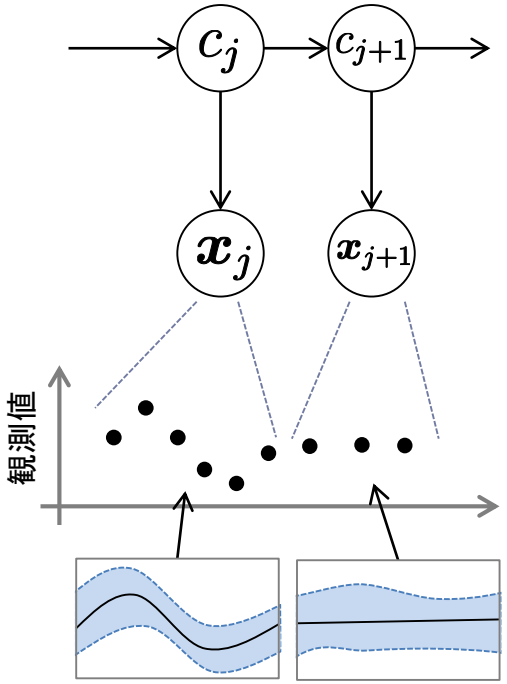


- ガウス過程で時系列情報の定型パターンを表現

GP-HSMMのパラメータ推定

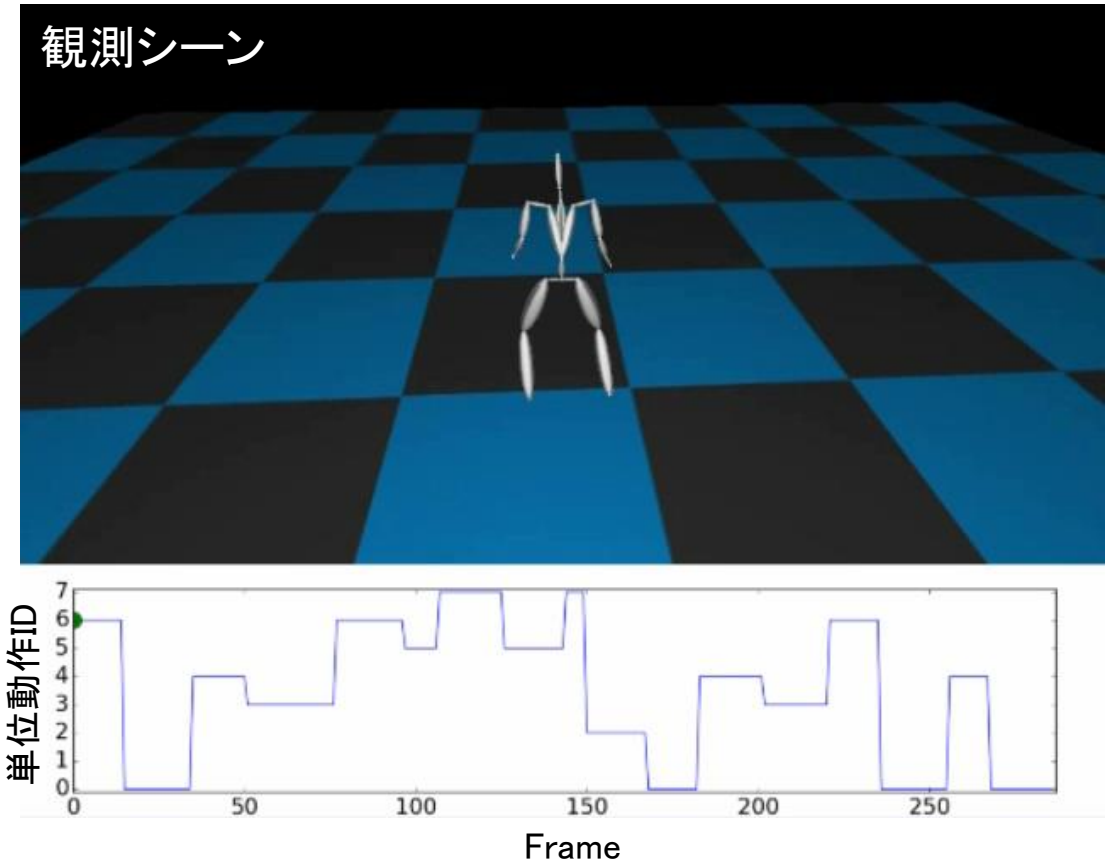
- HSMMでは1つの状態に分類される系列長は状態によって異なる
  ➡ 系列長も推定する必要がある
- パラメータ推定
  - 時刻tのデータ点を終点とした長さkの系列のクラスがcである確率からサンプリングすることで決定





  - あらゆるkとcの組み合わせの確率を計算する必要がある
    ＝動的計画法(Forward filtering-Backward sampling)を利用

モーションキャプチャの教師なし分節化

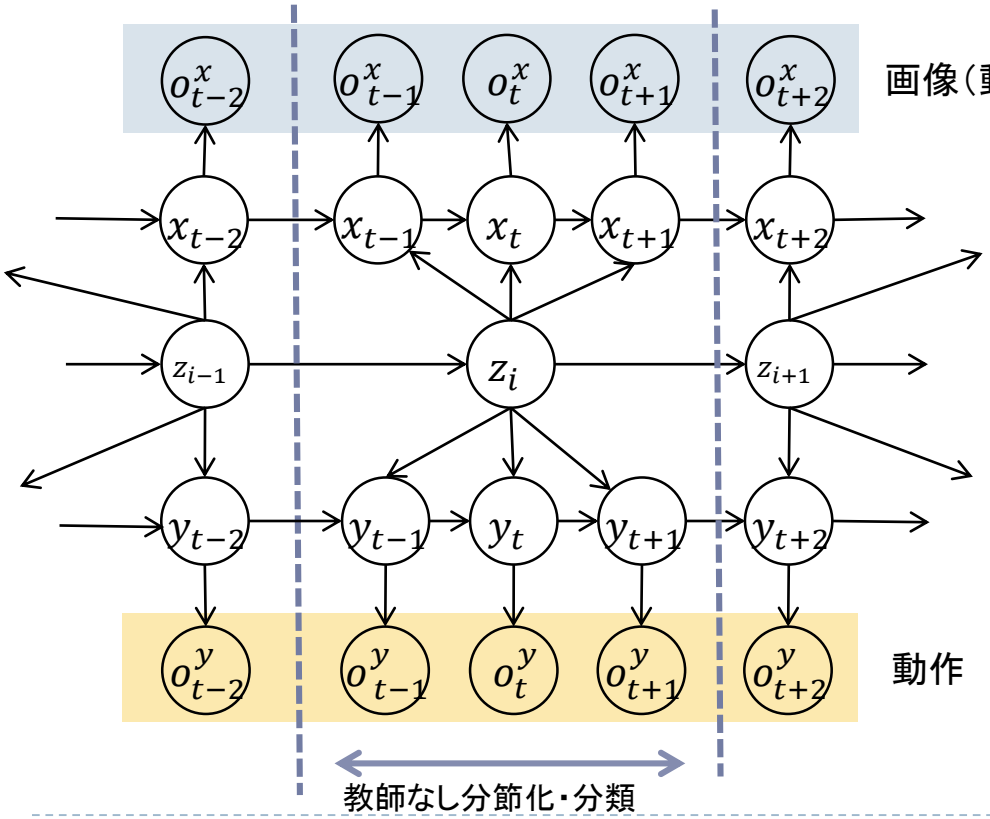


- 連続データを分節・分類することで単位動作の獲得ができている

現在の取り組み

マルチモーダル情報の分節・分類

- mMLDAの時間展開に基づく時系列マルチモーダル情報の分節・分類に基づく概念形成
  - どこからどこまでを1つの概念で表現可能化を教師なしで推定



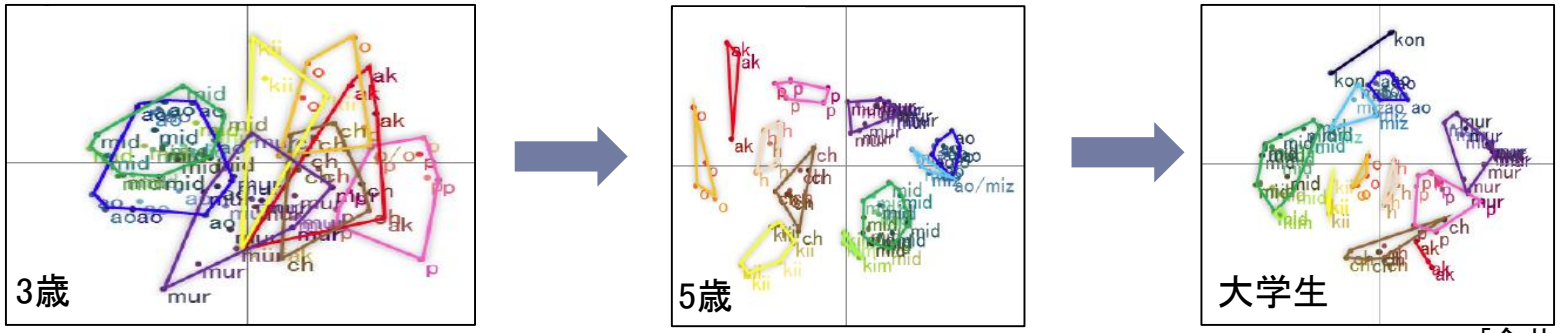
ロボットから取得可能な
連続センサ情報からの
柔軟な概念形成の実現

中村 他，"複数概念の時間的分節化に基づくロボットによる上位概念の学習"，人工知能学会全国大会，2017

人の概念獲得との比較

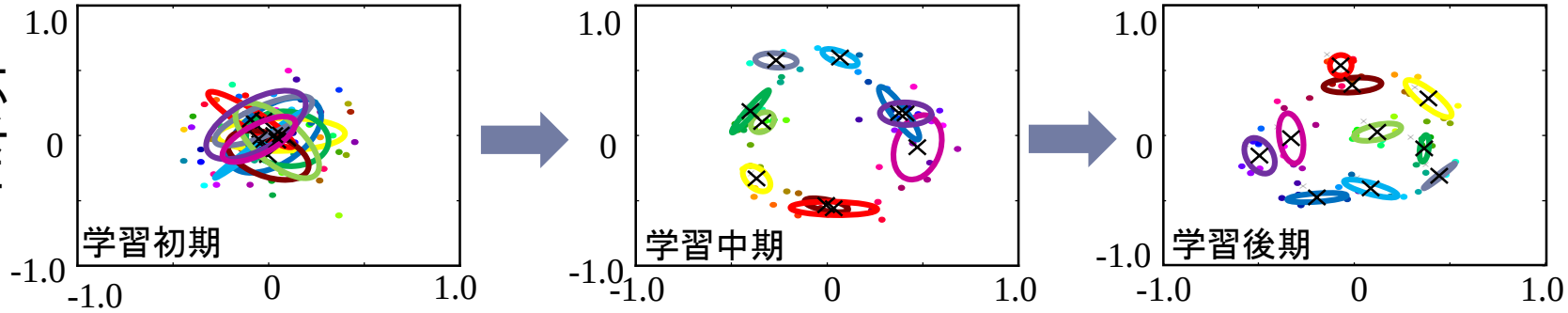
- 提案モデルによって概念獲得過程が再現可能か検証
  - 概念を二次元空間にプロット→似たような変化を確認
  - 教示発話の違いによる概念学習の変化の検証

人



[今井+ 2015]

ロボット



- 船田　他, "マルチモーダル概念形成における概念と言語の相互作用の解析", 人工知能学会全国大会, 2016
- Funada et al., "Analysis of the Effect of Infant-Directed Speech on Mutual Learning of Concepts and Language Based on MLDA and Unsupervised Word Segmentation", IROS2017: ML-HLCR, 2017

まとめ

まとめ

- 本発表ではロボットによる概念・語彙・語意・文法獲得に関する研究を紹介
  - ロボットが取得したマルチモーダルな情報から教師なしでボトムアップに概念を形成
  - 人の教示発話から語彙・文法を学習
    - 単語と概念を結びつけることで単語の意味の学習
    - 単語と結びついた概念の遷移規則を文法として学習
  - 獲得した言語の情報がトップダウンに作用することでより人の感覚に近い概念の形成が可能