

料理ロボットのための道具の選択・使用深層学習モデル - 道具と食材の配置に応じた料理のよそい動作の実現 -

斎藤 菜美子[†] 呉 雨恒[†] 尾形 哲也[‡] 森 裕紀^{*} 王 丹阳[†] 陽 品駒[‡] 菅野 重樹[†]

早稲田大学 創造理工学研究科 総合機械工学専攻[†] 早稲田大学 基幹工学研究科 表現工学専攻[‡]
早稲田大学 次世代ロボット機構^{*}

1. 序論

人間の生活空間で、多様な日常作業を支援することがロボットに望まれている。人間と同様、ロボットが道具を使用できれば、身体的制限を超えたタスク実行が可能となる。

現実世界において、道具の種類や使う環境、操作対象のオブジェクトは数多く、一つ一つ数値モデルを作成する[1]のは困難である。そこで、ロボット自らが状況に応じて道具を選択し、使いこなす手法が求められる。

2. 先行研究と本研究の目的

道具選択・使用の先行研究では、動作対象のオブジェクトは毎回同じ実験環境に設置されていた。そして、使うべき道具やロボットが行うべき動作は、実験者が与えた命令や最終目標時の画像に応じてロボットは判断していた[2]。

本研究では、オブジェクトや道具の設置背景や位置が変わっても対応でき、人間の補佐を必要とせず、その場の環境の認識結果のみに応じてロボットが行動決定できる学習モデルを構築することを目的とする。

ロボットの道具使用場面の一例として料理の場面を想定し、食材に応じておたま、またはフライ返しを選択して、鍋やフライパンから皿に食材を移すタスクを実現した。

3. 提案する深層学習モデル

提案モデルは図1の通り You Only Look Once (YOLO)[3], Convolutional Auto Encoder (CAE)[4], Multiple Timescales Recurrent Neural Network (MTRNN)[5]の3つの学習器から成る。

まずロボットを遠隔操縦し、複数の料理道具で、複数の食材を操作させる。その間取得したアー

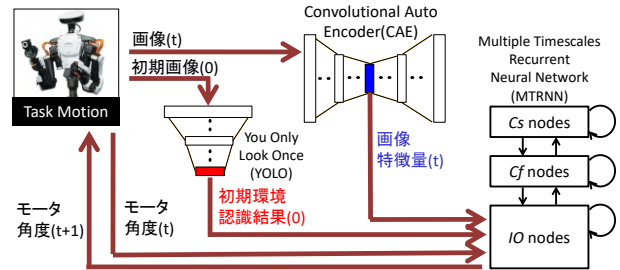


図1. 提案する道具選択・使用学習モデル

ムのモータ角度とカメラ画像の時系列情報を訓練データとする。提案モデルを用い、現在情報から次の時点の情報を予測できるように学習した。テスト時には、学習モデルに予測された関節角度に従ってロボットを動かすことを繰り返し、動作を生成させる。

3.1. YOLO

YOLOはConvolution layerを用いてクラス分別と bounding box の出力を同時に予測できる。そのため、物体検出処理を高速で行える。本研究では、検出性能の高いYOLOv3を採用する。

タスクの初期画像から、図2のように道具や食材の種類と初期配置位置を求める。そして、用いる道具と食材のクラスと bounding box の左上の角の座標を「初期環境認識結果」として出力する。

3.2. CAE

入力データがニューロン数の少ない中間層を経由後、再び元の次元数に出力される砂時計型の構造を有しており、出力データが入力データを復元するように学習される。これにより、画像を圧縮し、道具と食材、ロボットアームの特徴を、中間層から低次元の「画像特徴量」として取り出すことができる。

3.3. MTRNN

現在から次の状態を予測学習するRNNの中でも、異なる時定数を持つCf層、Cs層を有し、より複雑で長い時系列データを学習できる。初期環境

Deep Learning based Model for Cooking Robot to Select and Utilize Tool - Realization of serving motions according to the arrangement of tools and ingredients -

[†] Namiko Saito, Yuheng Wu, Danyang Wang and Shigeki Sugano with Department of Modern Mechanical Engineering, Waseda University

[‡] Tetsuya Ogata and Pin-Chu Yang with Department of Intermedia Art and Science, Waseda University

^{*}Hiroki Mori with Future robotics organization, Waseda University

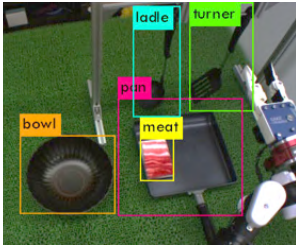


図 2. YOLO の出力

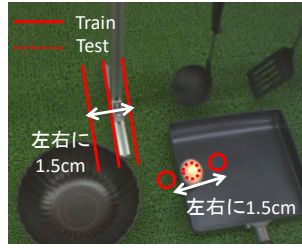


図 3. 初期環境設定

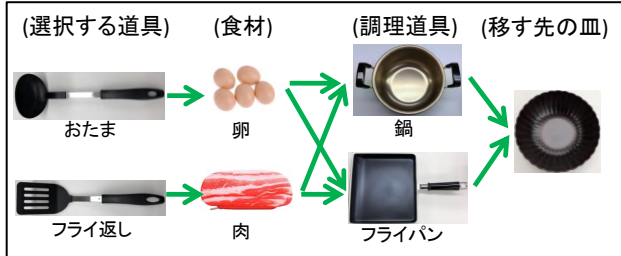


図 4. 使用する道具とオブジェクト

認識結果および、画像特徴量とモータ角度の時系列情報を統合学習する。この際 Cs 層の初期値 Cs(0)に一連のダイナミクス情報を保存する。これにより Cs(0)に道具や食材の特徴が自己組織化されることが期待される。

4. 実験設定

用いる道具と食材は、ロボットの目の前の机の上に設置し、初期位置は図3のように設定する。

ヒューマノイドロボット NEXTAGE を使い、図4に示す道具と食材の組み合わせで、調理道具内から皿にオブジェクトを移動させる。

取得するデータは、ロボットの右アームの関節角度(6 自由度)とグリッパーの角度、および右目に搭載されたカメラの画像である。計 24(道具の種類(2) x 調理道具(2) x 食材位置(2) x 道具位置(2))の時系列データを予め取得し、学習モデルの訓練に用いた。

5. 結果

図3の点線で示した、訓練していない位置に食材を置いた際も、人の助けを必要とせず、道具を選択・使用する動作が生成された(図5)。

また、図6に、Cs(0)の値を主成分分析し2次元にした結果を示す。道具と食材が自己組織化されていることが確認できた。

6. まとめ

道具や環境のモデリングを必要とせず、ロボット自らが使うべき道具を選択・使用する学習モデルを構築した。未学習位置に食材を配置した際も動作生成でき、環境設定に応じて行動を適応できることを示した。

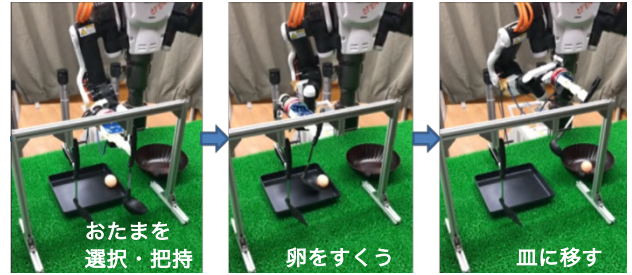


図 5. 未学習位置における動作生成の様子

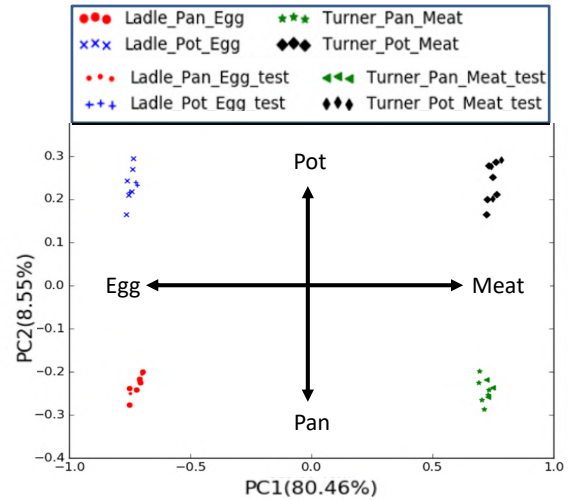


図 6. Cs(0)の主成分分析結果

謝辞

本研究成果の一部は、文部科学省科研費基盤研究 (S) (No. 25220005), (A) (No. 19H01130), JST CREST (No: JPMJCR15E3) の助成を受けたものです。本研究成果の一部は、早稲田大学理工研プロジェクト研究「自然と共生する知能情報機械系に関する基盤研究」の一環として行われたものです。ここに謝意を表します

参考文献

- [1] G. Pan, F. Yang and M. Chen, "Kinematic control of a dual-arm humanoid mobile cooking robot" In i-CREATE 2018 Proceedings of the 12th International Convention on Rehabilitation Engineering and Assistive Technology, pp. 308-311, 2018.
- [2] N. Saito, K. Kim, S. Murata, T. Ogata and S. Sugano, "Tool-Use Model Considering Tool Selection by a Robot Using Deep Learning," 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids), pp. 270-276, 2018.
- [3] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.
- [4] J. Masci, U.Meier, D.Ciresan, and J. Schmidhuber, "Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction," LNCS 6791, pp.52-59, 2011.
- [5] Y. Yamashita, J. Tani, "Emergence of Functional Hierarchy in a Multiple Timescales Recurrent Neural Network Model: A Humanoid Robot Experiment", PLoS Comput Biol, 2008.