

物体検出を用いた調理の時系列パターンによる分類

Clustering time-series data of cooking using object detection

森 梓^{*1} 濱 龍太郎^{*1} 原田 篤^{*1} 高久 由香里^{*1} 橋本 敦史^{*2}
 Azusa Mori Ryutaro Hama Atsushi Harada Yukari Takaku Atsushi Hashimoto

^{*1} 株式会社 LIXIL ^{*2} 京都大学教育学研究科
 LIXIL Corporation Kyoto University

This paper presents an appliance of object detection with deep learning for a series of cooking video frames and an analysis of these time-series detection data. Our final goal is offering personalized kitchens. We have observed and analyzed cooking behaviors with visual observation and spread sheet software. However this conventional method was too tough to perform with sufficient precision and subject numbers. Furthermore, it was difficult to analyze from various viewpoints. So, we introduced object detection with deep learning as a tool for automation of the method. We applied object detection for a series of cooking video frames which contains about 90 objects, extracted these data and analyzed with Python, and found that the time-series data counting number of detected objects clustered 3 patterns in sink area. This discovery contributes to understanding of cooking patterns which leads to offering personalized kitchens.

1. はじめに

キッチン毎日、何回も、長年にわたって使い続けるため、使いやすさが重要である。これまで弊社では行動観察やアンケート、インタビュー等の調査手法を用いて、ユーザーの潜在的な困りごとを発見し、新しいキッチンの提案につなげてきた[村上 2010]。

これらの調査手法の中でも行動観察※[松波 2013]は、得られる情報が多い反面、工数が膨大にかかる点が難点であった。一般的な主菜、副菜、汁物、ご飯から構成されるメニューの調理には約 1 時間かかる。この 1 時間の調理行動をキッチンの天井に取り付けたカメラで撮影し、動作単位に分解する。動作はのべ約 1000 回含まれ、そこに時間(秒)、座標(ブロック単位)、調理アイテムの種類(調理道具・食材約 150 種類)、調理アイテムの状態、設備(水栓、コンロ、収納等)へのアクセス、キッチンスペースの使われ方等の情報を付随させる。このような分析を目視観察と Excel をベースにした従来のスタイルで行う場合、1 件あたり約 3 か月かかった。工数的にも分析者のモチベーションにも限界があり、他にも、一定の品質での精密なデータが取得できない、母数を増やせない、集計分析にとどまる等の問題があった。然して、分析視点も限定的にならざるを得なかった。

そこで私たちは調理アイテムを自動検出させることを考え、深層学習による物体検出技術を採用した。理由は 2 つある。ひとつは、特定のレシピに応じた学習サンプルさえ用意すれば食材の色や形の変化、分裂や結合にも対応可能であること、もうひとつは、将来的に実験用キッチン以外でのデータ取得の可能性を配慮し、カメラひとつで対応可能であることである。この技術を用いることにより一定の品質かつ精密なデータで母数を多く取得できれば、集計分析だけでなく統計分析、さらには機械学習

による分析も可能となる。そうすることで分析・発見の切り口を増やし、より使いやすいキッチンを一人一人にフィットした形で提案を行っていきたいと考えている。

今回はファーストステップとして、調理動画に深層学習による物体検出技術を適用し、そこから得られたデータをもとに、キッチン上に出現する調理アイテム数の時系列変化について分析した結果を報告する。

2. 目的

キッチンのパーソナライズ提案につなげるために、調理動画に深層学習による物体検出技術を適用して、より一定の品質かつ精密なデータで母数を多く取得する。これにより分析・発見の切り口を増やして、調理のユーザー特性を把握する。

3. 目標

本報告では、キッチンにおける調理のユーザー特性として、調理アイテム検出数の時系列変化を取り上げる。24 名の調理動画に深層学習による物体検出技術を適用して、89 種類の調理アイテム(食材・調味料・調理道具・食器)、時刻、座標のデータを取得する。ここからエリア別に調理道具・食器のデータを抽出し、調理アイテム検出数の時系列変化に基づく分類を行う。分類した結果はもとの調理動画ならびにアンケート・インタビューと照合し、定性的な妥当性を確認する。最後に、各分類に対して提案されるキッチンのイメージを取り上げる。

4. 実験

4.1 実験条件

4.1.1 環境

調理モニター調査は弊社施設内の実験環境にて行った。実験環境全体図を図 1 に示す。LIXIL 製キッチン RICHEL SI (幅:2585 mm、奥行:970 mm(650 mm)、左勝手;左手にシンク)を使用した。シンク横スペース、水切りカゴ、シンク、作業スペース、コンロの横幅は、それぞれ 250 mm、230 mm、540 mm、750 mm、815 mm である。なお、本キッチンはカウンタータイプであるが、より一般的な壁付けタイプを想定して奥行 650 mm 以上のエリアは使用しないこととした。

連絡先: 森 梓, 株式会社 LIXIL, Technology Research 本部
 人間情報科学研究所, 〒136-8535 東京都江東区大島 2-1-1 LIXIL WING ビル, azusa.mori@lixil.com

※ここでは狭義の行動観察、すなわち行動観察調査における発見フェーズをさすものとする。本来、行動観察調査という言葉は、①発見する(事実の収集)、②再構築する(気付きと考察)、③着想を得る(解釈からの発想)、④想像する(経験のデザイン)の全体をさす[松波 2013]。

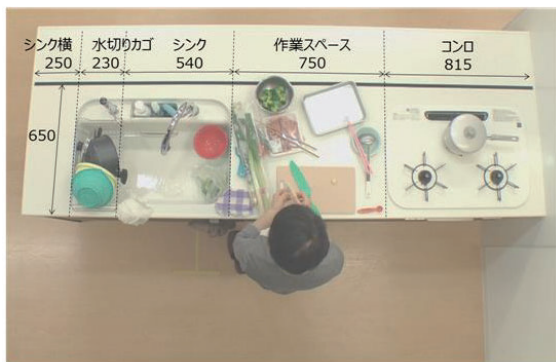


図1. 実験環境全体図

撮影は天井に設置したカメラ(Panasonic 製 AG-HMR10A)を用いて 29.97 fps で行い、1920×1080 pixel で保存した。

4.1.2 調理モニター(動画の取得)

被験者は、一般、社員含めて 20 代～60 代の男女 24 名とした。調理メニューは、調理の基本動作のうち「切る、和える、味付け、約、ゆでる、盛り付け」を含むように設計し、照り焼きチキン(鶏もも肉 2 枚)、添え野菜(ゆでブロッコリー、カットトマト)、酢の物(きゅうり、しらす)、味噌汁(大根、しめじ、ねぎ)、ごはん(2 合; 電気炊飯器使用)とした。食材 2 名分とその他の調理アイテム(調味料・調理器具・食器)はあらかじめ用意した。調理メニューは被験者に事前告知し、用意した調理アイテムの収納場所や機器の使い方以外の調理作業・動作の教示は行わず、被験者それぞれの手順で調理させた。なお、調理された料理は被験者と実験者で摂食し、あわせて 10 項目のアンケートとインタビューも実施した。アンケート内容は自宅キッチンでの調理に関するもので 5 段階評価とした。

4.2 予備調査

調理におけるユーザー特性を把握するために、本報告では調理アイテム検出数の時系列変化に着目する。以下の本調査に先立ち、上記 24 名外の 2 名の調理動画について、Excel を用いて 5 秒単位で分解し、調理アイテム検出数の時系列変化に明確な差異があることを確認した(図 2)。この予備調査により、調理アイテムの出し入れや洗いのタイミングに差があることが示唆された。

4.3 深層学習による物体検出

4.3.1 データセット

被験者 24 名中 16 名の調理動画から 0.5fps でフレームを抜き出し、33,405 枚の画像とし、出現した調理アイテム(食材・調味料・調理器具・食器)89 種類すべてにアノテーションしたものをデータセットとした。

4.3.2 SSD の学習・検出

SSD512 に ResNet50 の事前学習済みのモデルを適用し学習を行った[Liu 2016]。学習・検出用の画像は全て 512×512 pixel に縮小して処理を行った。PASCAL VOC 基準で評価を行うと mAP (mean Average Precision)は 0.62 であった。調理アイテムの中で出現回数が少なかったり、形状が大きく変化したりする食材や調味料の検出精度が低かった。このモデルに 24 名の調理動画を供した。検出精度については今後の課題としたい。

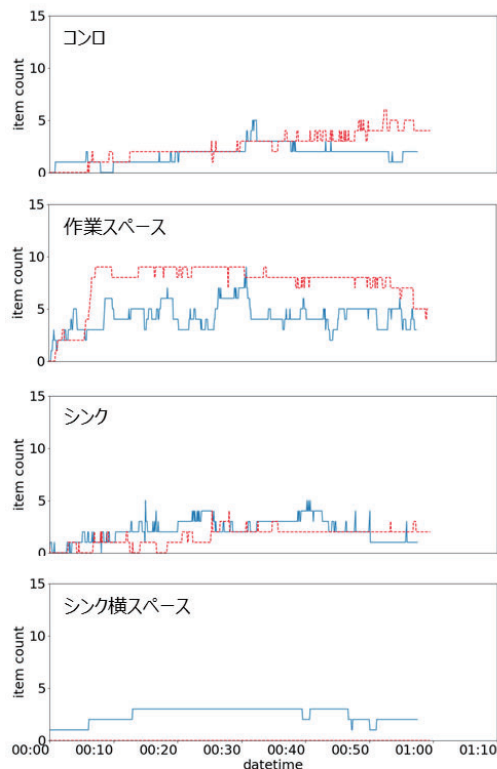


図2. 調理アイテム数の時系列変化(目視・手動記録; n=2; — 被験者A, --- 被験者B)

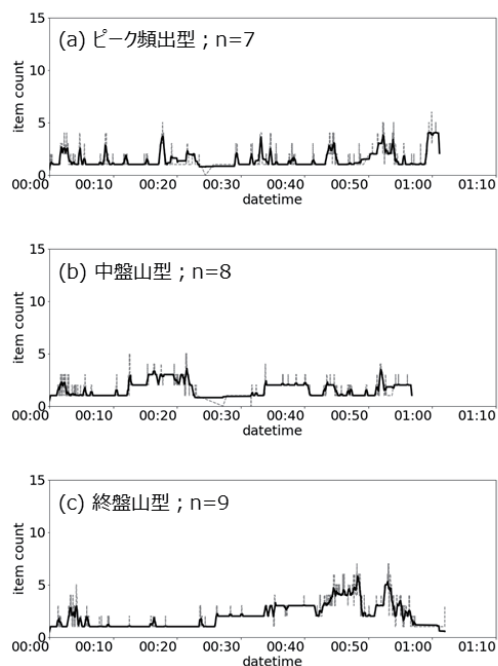


図3. シンクエリアにおける調理アイテム検出数の時系列変化のパターン分類(機械学習による物体検出ならびに目視分類; n=24; --- 原系列, — 移動平均(n=60))

4.4 データ抽出

上記により取得した調理アイテム、時刻、座標の 3 つのデータから、キッチン全体、シンク横スペース、水切りカゴ、シンク、作業スペース、コンロの 6 エリアについて、それぞれ調理アイテム検出数の時系列変化を抽出した。本報告では、検出精度の低い食材・調味料は検出数に対する影響が大きいため除去し、調理道具・食器の検出数をカウントした。なお、抽出にあたっては、解析環境の制約により、1/10 にダウンサンプリング処理を施した。

4.5 調理アイテム検出数の時系列変化

4.5.1 各エリアにおけるユーザー特性

6 エリアでの調理アイテム検出数の時系列変化について、目視分類を行った。この際、判別のしやすさのために、移動平均 (n = 60) に変換した。
本報告では、予備調査で示唆された洗いのタイミングに関係のあるシンクエリアに着目することとした。他のエリアの分類妥当性や、エリア間のユーザー特性の関連性、あるいは分類の自動化 (統計処理あるいは機械学習) については今後の課題とする。

4.5.2 シンクエリアにおけるユーザー特性

シンクエリアに着目すると、調理アイテム検出数の時系列変化のパターンは 3 つに大別された。ピーク頻出型、中盤山型、終盤山型の 3 つである。これら 3 つのパターンの代表的な例を図 3 に示した。点線は原系列、実線は移動平均である。
ピーク頻出型では細切れに短時間のピークが現れている。他方、中盤山型や終盤山型では、検出数が増加した状態が一定時間以上継続する “山” が現れる。この山の出現タイミングが調理中盤にあるか (終盤にも現れるパターンを含む)、調理終盤のみにあるかで判別を行った。
なお、ここで出現するピークあるいは山は、後述するように、「洗い」動作のみではないことに注意が必要である。

4.6 考察

4.6.1 妥当性検証 (動画照合)

分類の妥当性を検証するために、被験者 24 名中 8 名のデータについて、以下の 3 つの作業を行った。まず、調理アイテム検出数の時系列変化に現れたピークを抽出した。次に、ピーク時刻における動作を、調理動画から目視で手動記録し、動作を「洗う、拭く、出し入れ・移動、注ぐ・流す、左手の映り込み、その他」の 6 種類に区分し、被験者毎に動作種類の割合を算出した。ここで、「左手の映り込み」とは、作業エリアにおいて右手で何らかの動作をしているときに、左手に把持した調理アイテムがシンクエリアに映り込んでいる場合を指す。また、「その他」はシンクエリアでの動作が見られない場合を指す。この原因については調査中であるが、検出精度 (検出逃しや誤り) の時系列でのばらつきが、ピークとして抽出されている可能性が高い。最後にこれを分類結果と照合した。照合結果を図 4 に示す。
調理工程のリセットにかかわる動作、すなわち「洗う、拭く、出し入れ・移動」の割合を比較すると、ピーク頻出型では平均 56.7%、中盤山型と終盤山型では平均 41.7%となった。この差について有意水準 5% で t 検定を行ったところ、 $t(6) = 2.83$, $p = .03$ であり、有意であると認められた。ピーク型と山型の 2 種の判定は、調理工程の前進に必須ではないリセット動作の割合の多寡と一致し、妥当であるといえる。

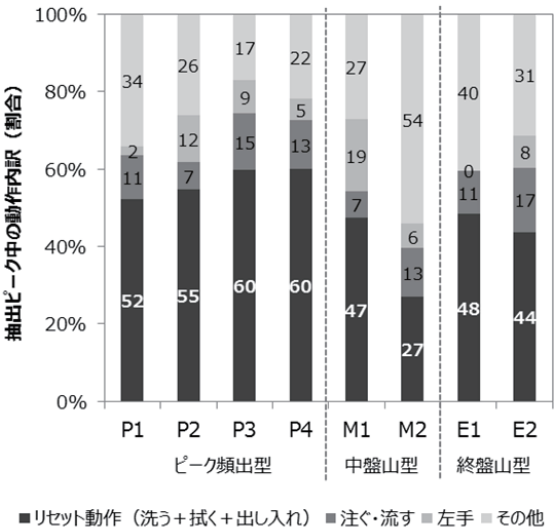


図4. 抽出ピーク中の動作内訳と3分類の照合 (n=8)

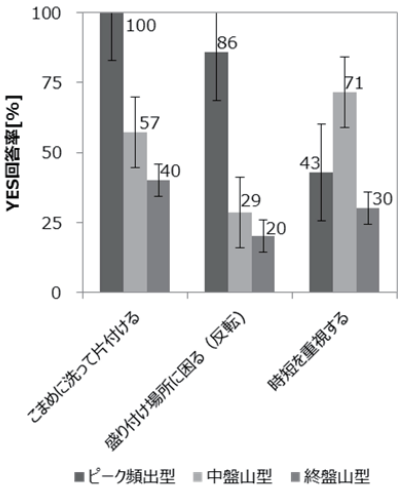


図5. アンケート回答と3分類の照合 (n=24)

表1. 3つの分類型における調理時間 (中央値; n=24)

	ピーク頻出型	中盤山型	終盤山型
調理時間	1:10:47	1:04:41	1:10:40
片付け時間	0:19:05	0:21:58	0:21:51
合計時間	1:27:50	1:25:18	1:31:48
Δ調理時間	0:06:06	0:00:00	0:05:59
Δ片付け時間	- 0:02:53	0:00:00	- 0:00:07
Δ合計時間	0:02:32	0:00:00	0:06:30

ここで、さらに、リセット動作の割合に差が生じる原因を探るために、各分類型のリセット動作自体を比較した。リセット動作のひとつである「まな板洗い」に着目すると、ピーク頻出型では、例えばブロッコリーを切り終わった時点で即座に洗う等、各メニューの各工程が完了するたびに発生した。中盤山型では、フライパンにひいた油を温めている間に、フライパンの状態を確認しながら洗い、油が温まると鶏もも肉をフライパンに載せ、焼き加減を見ながら再度洗いに戻るといように、待ち時間にまな板洗いが発生した。終盤山型では、まな板洗いは発生せず、野菜を切っていたまな板で鶏もも肉を切ると、その後は新しいまな板が

表2. キッチンを対象とする行動観察の量的進化

	第1世代	第2世代	第3世代
手法	目視・手動記録	目視・手動記録	自動検出・自動記録
可視化形態	図化	疎な数値化	密な数値化
N数	1件	2件	制限なし（今回24件）
分析	（センスと経験）	・集計分析 ・比較分析	・集計分析 ・統計分析 ・機械学習分析
記録	・調理動作（約1000回） ・時間（秒） ・調理アイテムの種類（約150種類） ・設備（水栓、コンロ、収納等）へのアクセス ・手の状態（ドライ、ウェット）	・調理動作（約1000回） ・時間（秒；動作発生毎） ・座標（25cmブロック単位） ・調理アイテムの種類（約150種類） ・調理アイテムの状態（使用・待機・放置・完了） ・設備（水栓、コンロ、収納等）へのアクセス	・調理アイテムの種類（今回89種類） ・時間（ミリ秒；29.97fps） ・座標（1cm単位）
手動計算	－	・調理アイテム数の時系列変化 ・ブロック毎のアイテム出現特性 他	－
自動計算	－	－	・調理アイテムの時系列変化 ・エリア毎の調理アイテム出現特性 ・調理アイテムの状態（移動・滞留）他
主なツール	・Power Point（可視化）	・Excel（計算、可視化）	・深層学習（物体検出） ・Python（抽出、計算、可視化） ・Tableau（可視化）

取り出された。このことから、山型の 2 種類の分類も妥当であると考えられる。

4.6.2 妥当性検証(アンケート照合)

他方、自宅での調理に関するアンケート 10 項目 (5 段階評価)との照合も実施した。上述の 3 分類と合致する傾向が見られた質問項目は、「こまめに洗って片付けるか」、「盛り付け場所に困るか」であった。照合結果を図 5 に示す。「こまめに洗って片付けるか」という自意識に対する YES 回答率は、ピーク頻出型では 100%であったのに対し、中盤山型と終盤山型では 50%程度にとどまった。また、「盛り付け場所に困るか(反転)」に対する YES 回答率は、ピーク頻出型では 86%、中盤山型と終盤山型では約 20%であった。これらの照合から、ピーク型と山型 2 種の分類判定は、被験者の自意識とも合致していた。

なお、インタビューにける「時短を重視するか(YES、NO、回答なし)」とも照合を行ったところ、YES 回答率は、ピーク頻出型 41%、中盤山型 71%、終盤山型 30%と、明確な差が出た(図 5)。特筆すべきは、中盤山型の割合の高さと、終盤山型の積極的に NO という被験者の存在である。

そこでさらに調理時間実測値と照合してみると、時短重視の中盤山型は、ピーク頻出型と終盤山型に比べて、約 6 分早いことがわかった(表 1)。調理中盤に現れるリセット動作は、時短意識に基づいていると考えられる。

終盤山型の NO と回答した被験者のインタビューからは、「追い込まれたくない」、「気ままに調理を楽しみたい」といった意見を聞くことができた。これらの被験者は、調理において便利さやスピードといった合理性を追求していない。このような特徴的な価値観は、行動観察だけでなくアンケートやインタビューとの掛け合わせから気付きを得る場合が多い。私たちはユーザーがそれぞれ求めることを、行動データの解析と傾聴の双方から理解し、パーソナライズ提案につなげていきたいと考えている。

4.6.3 各分類に対する提案

以上 2 つの検証から、シンクエリアにおけるユーザー特性の 3 分類は妥当であるといえる。これらの各分類に対し、例えば、ピーク頻出型にはこまめにリセット動作が行いやすいキッチン、中盤山型にはリセット動作と確認動作を並行しやすいキッチン、

終盤山型には盛り付けまで広々と使いやすいキッチン等、調理のユーザー特性にフィットした提案が可能である。

5. まとめと今後の展開

調理動画に深層学習による物体検出技術を適用した。この効果は表 2 の通りであり、工数削減が 100 分の 1、被験者数増加が 12 倍(今回)、粒度細密化が時間 10 倍、空間 25 倍、さらに均質性向上、すなわち同一のデータを用いて複数名が別に異なる視点で分析可能となった。

得られたデータから、6 エリアにおける調理アイテム検出数の時系列変化を分析した。この中で、シンクエリアにおけるユーザー特性を目視分類すると、ピーク頻出型、中盤山型、終盤山型の 3 つに大別された。これらの分類について、動画ならびにアンケートと照合すると、妥当性は良好であるといえた。

各分類に対しては、例えば、ピーク頻出型にはこまめにリセット動作が行いやすいキッチン、中盤山型にはリセット動作と確認動作を並行しやすいキッチン、終盤山型には盛り付けまで広々と使いやすいキッチン等、調理のユーザー特性にフィットした提案が可能である。

今後は、検出精度の向上、統計処理や機械学習を用いた自動分類、洗い等の動作の自動判定、シンクエリア以外のエリアにおけるユーザー特性との掛け合わせ、あるいはアイテム検出数の時系列変化以外の切り口での分析ならびにその結果との掛け合わせ等の課題に取り組み、より確度の高いパーソナライズ提案につなげていきたい。

参考文献

[村上 2010] 村上孝子: システムキッチンの調理スペースについての一考察～調理補助アイテムの提案～, 人間生活工学, vol.11, no.1, 通巻 33 号, 2010

[松波 2013] 松波晴人: 「行動観察」の基本, ダイヤモンド社, 2013

[Liu 2016] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., and Berg, A.: Ssd: Single shot multibox detector, in European conference on computer vision, pp. 21–37, Springer (2016)