手首装着型の加速度センサを用いた実時間調理行動認識手法

Real-time cooking activity recognition using accelerometers

大神 健司 [1*] 飛田 博章 [1]
Kenji Ogami[1*] Hiroaki Tobita[1]

[1] 東京都立産業技術大学院大学 Advanced Institute of Industrial Technology
[*]Corresponding author: Kenji Ogami, a2109ko@aiit.ac.jp

Abstract Although cooking by ourselves is important to manage our health, it is difficult especially for beginners to reproduce recipes. As a solution to this issue, systems supporting persons who are cooking by feedback from their activities are considerable. In this study, we propose a method of real-time cooking activity recognition using accelerometers to realize such systems. Our proposed method classifies the subject’s activities during cooking of curry to 7 classes by convolutional LSTM with frequency of one time per a second. We confirmed a performance of our proposed method was 86.39% in batch inference, 61.32% in real-time inference as results of evaluation with macro-average of recalls.

Keywords cooking recognition; wearable sensor; accelerometer

1 はじめに

自分でメニューを考え料理する自炊は，好きなものを選びがちな外食に比べ，健康を維持するために効果的である．しかし，料理初心者にはレシピの再現自体が難しく，レシピに記されている調理行動を再現できないため，味や見た目がレシピと大きく異なる結果となる．この課題の解決策として，料理初心者の行動に対し，次の調理工程の指示や，指示と異なる行動をした場合にそれを正すためのフィードバックの実現が重要となる（図 1）．

本研究では，調理者の行動を認識し，システムが適切にフィードバックする仕組みの実現に必要となる，実時間での調理行動認識手法について述べる．提案手法では，実際の調理で両手首にウェアラブル型の加速度センサを装着し，加速度と角速度のデータ（以後，両者を合わせて加速度データと表記する）を取得する．取得したデータに対し，調理行動を 7 クラス（「手を洗う」，「皮を剥く」，「切る」，「みじん切りにする」，「炒める」，「混ぜる」と，「その他」）にラベル付けし，畳み込み LSTM で学習・分類する．作成済のテストデータに対して推論を行い（本稿では，以後バッチ推論と表記する）学習済モデルの性能を評価した結果，再現率のマクロ平均が 86.39%であった．また，加速度センサを装着した状態で実際にカレー（図 2）を調理しながら取得したデータを実時間で処理して推論を行い（本稿では，以後実時間推論と表記する）性能を評価した結果，同マクロ平均が 61.32%であった．

2 関連研究

近年，調理をサポートするため，様々なセンサから得たデータを基に調理者の行動を認識する手法が多数提案されている．[1]，[2]，[3] では，調理者の身体に装着したウェアラブルカメラで一人称動画を撮影し，画像認識により調理者の調理行動を認識する手法を提案しているが，ウェアラブルカメラの向きや装着位置によっては認識対象の動作が画角に収まらないことや，プライバシーへの懸念から日常生活における動画撮影を忌避する意見も多い．[4] では，調理音から調理者の調理行動を認識しているが，音響信号には環境音等の雑音が混入し易く，安定した品質のデータ取得に課題がある．[5]，[6] では，離れた位置に設置した深度センサで調理者の身体を三次元的に捉え，三次元空間内における調理者の身体の動きから調理行動を認識する手法を提案しているが，深度センサという特殊な設備が必要であるため，コストに課題がある．[7]，[8] では，加速度センサを包丁の柄に取り付け，それぞれ調理者の調理技術を向上させる手法と，動画レシピと実際の調理工程を同期させる手法を提案しているが，包丁に加速度センサを取り付けているため，包丁を使用する調理工程以外には適用できない．[9] では，調理者の身体に装着した複数の加速度センサとモーションキャプチャにより得られたデータを利用し，複数の参加者が提案した手法で認識精度を競った結果を総括している．しかし，提案された手法は取得済みのデータを対象とした手法に限っており，調理者の調理行動を実時間で認識してサポートすることを考慮していない．

図 1 システムデザイン

3 データセットの作成

モデルを学習するための学習データとテストデータは，単一の被験者が別の日に加速度センサを装着した状態で，実際にカレーを調理して取得したデータから作成した．両手首の加速度センサの動作クロックが非同期であるため，得られたデータをサンプリング周波数変換処理で同期化した上でデータセットを作成している．

3.1 加速度データの取得

両手首に装着したウェアラブル型の加速度センサ，以後加速度センサと表記する）により，70Hz のサンプリング周波数で加速度データをサンプリングする（図 3）．

加速度センサと PC は Bluetooth で通信し，通信の開始時に PC のシステム時刻に加速度センサの内蔵リアルタイムクロックの時

Received: 2024-10-30; Published online: 2025-02-28

刻を合わせる．リアルタイムクロックの時刻は，サンプリングした加速度データにタイムスタンプとして付与する．

両手首に加速度センサを装着した状態で実際にカレーを調理し，カレーの全調理工程を通して加速度データを Bluetooth 経由で PC に取得する．

3.2 加速度データと実際の調理行動との対応

加速度データの取得と合わせ，被験者の額に装着したウェアラブルカメラで一人称動画を撮影する．動画にはタイムスタンプを表示しておき，取得した加速度データのタイムスタンプと照合することで，加速度データと実際の調理行動との対応が識別できるようにした．

図 2　カレーの調理

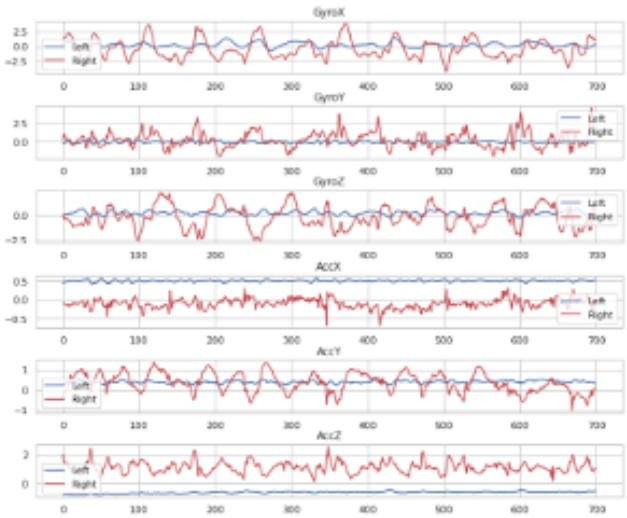


図 3　センサデータ

3.3 データの構造

モデルに入力する加速度データの構造は，両手首の加速度（X，Y，Z）と角速度（X，Y，Z）の合計 12 軸のデータ 10 秒間分（700 サンプル）とする．

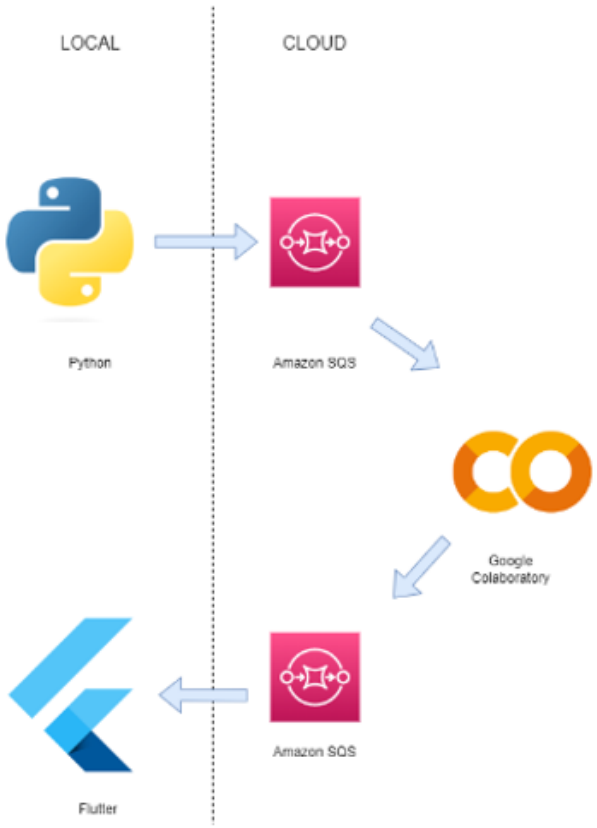


図 4　システム実装

3.4 データの同期化

両手首に装着した加速度センサは，PC との通信開始時に PC のシステム時刻を取得し，内臓リアルタイムクロックの時刻を PC に合わせている．しかし，両手首の加速度センサ同士の動作クロックが非同期であるため，加速度データのサンプリング周期に誤差が生じる．これにより，例えば 1 秒間サンプリングして得られる加速度データのサンプル数が左右で異なる結果となる．サンプリング周期の誤差は固定的であるため，長時間サンプルする程サンプル数の差は累積するが，単一のモデルで左右の加速度データを取り扱う場合，左右のサンプル数は一致している必要があり，サンプル数の差を解消しなければならない．

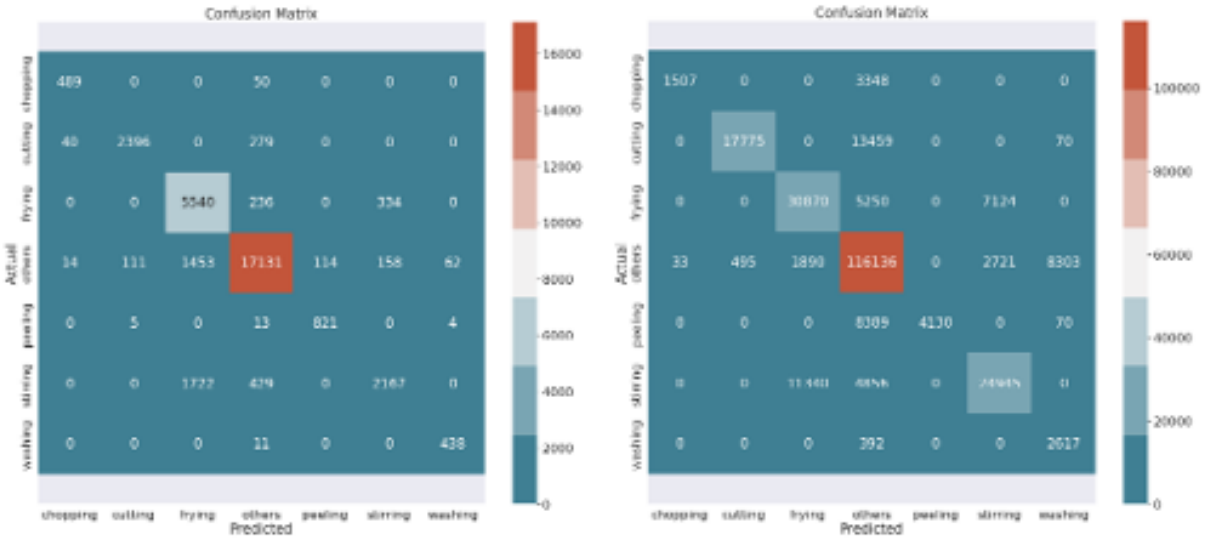
提案手法では，右手首から得られる加速度データを基準とし，左手首の加速度データを 10 倍のサンプリング周波数（700Hz）に一旦変換してタイムスタンプを付与し直し，右手首の加速度データのタイムスタンプと時間的に最も近いサンプルを左手首の加速度データとして再サンプリングする．

3.5 正規化

カレーの全調理工程から得られた加速度データを対象として，ロバストスケーラで正規化を行う．

3.6 ラベル付け

提案手法では，カレーの調理工程を，「手を洗う」，「切る」，「皮を剥く」，「みじん切りにする」，「炒める」，「混ぜる」，「その他」にクラス分けする．最初に，取得した加速度データ全体に「その他」のラベルを付ける．その上で，加速度データの取得と同時に撮影した一人称動画の中に「その他」以外の 6 クラスの動作を見付ける．動画内のタイムスタンプと加速度データのタイムスタンプか



クラス名	再現率 (%)	マクロ平均 (%)
手を洗う	97.55	86.39
切る	88.25	
皮を剥く	97.39	
みじん切りにする	90.72	
炒める	90.67	
混ぜる	50.19	
その他	89.96	

クラス名	再現率 (%)	マクロ平均 (%)
手を洗う	86.97	61.32
切る	56.78	
皮を剥く	32.81	
みじん切りにする	31.04	
炒める	71.39	
混ぜる	60.63	
その他	89.63	

図 5　バッチ推論と実時間推論の結果と性能

ら，見付けた動作に対応する加速度データの時間軸上の位置を特定し，適切なラベルに付け替える．この作業により全ての「手を洗う」,「切る」,「みじん切りにする」,「皮を剥く」,「炒める」,「混ぜる」の 6 動作にラベルを付けると，6 動作以外の全ての動作が「その他」にラベル付けされたまま残る．

4　モデル

調理行動を認識するための深層学習のモデルには，時系列データの処理に適した LSTM と GRU に加え，[10] で提案している畳み込み LSTM の 3 種類を実際に学習して評価し，最も性能の良かった畳み込み LSTM を採用した．

5　実時間調理行動認識の実装

提案手法の実時間推論では，PC に取得した加速度データを計算リソースの豊富なクラウドにアップロードして処理を行う．実装上ボトルネックであったデータの同期化処理を並列化することにより，1 秒間に 1 回の推論頻度を実現した．

5.1 システム構成

両手首の加速度センサから Bluetooth 経由で加速度データを PC に取得し，1 秒間に 1 回の頻度で 1 秒間分の加速度データをクラウドにアップロードして実時間推論を実行する．クラウドで実時間推論を実行後，PC に分類した結果を転送する（図 4）．

5.2 並列処理

実時間推論の処理では，1 秒間に 1 回の頻度で 1 秒間分のデータ（約 70 サンプル）を受信し，過去の 9 秒間分のデータと合わせて処理を実行する必要がある．GPU を活用することのできないデータの同期化処理が実装上のボトルネックであったため，処理を並列化することでボトルネックを解消した．具体的には，加速度データをアップコンバートしてタイムスタンプを付与する処理を 8 並列化し，左手首の加速度データを再サンプリングする処理を 4 並列化した．

6　評価

提案手法による学習済モデルと，実時間推論の性能を評価した結果を示す（図 5）．

6.1 実時間推論の評価方法

実時間推論の性能評価に必要なデータの取得は，データセットの作成時と同一の被験者 1 名が両手首に加速度センサと額にウェアラブルカメラを装着した状態で，実際にカレーを調理しながら実時間推論の実行と，一人称動画の撮影を行って取得した．実時間推論の結果は，加速度データの各サンプルに分類の結果として付与し，調理後に一人称動画を確認して学習データの作成時と同じ要領で正解ラベルも加速度データに付与した．得られた実時間推論の結果と，正解ラベルより各クラスの再現率と，それらのマクロ平均を得た．

6.2 評価結果に対する考察

実時間推論の性能がバッチ推論の性能に対し，著しく低い結果となった．提案手法では 10 秒間のウインドウを 1 秒ずつシフトさせて調理行動を分類しており，例えば 10 秒間のウインドウの中に「切る」動作が 1 回だけ入っているデータでは，「切る」動作がウインドウの何処にあっても分類できるように学習データを作成すべきところ，「切る」動作がウインドウの中心に位置するデータば

かりを使用してモデルを学習したため，「切る」動作がウインドウの端にあるデータを「切る」と分類できなくなったこと等が実時間推論の性能を劣化させた要因として挙げられる．

7 まとめ

本研究では，調理者の行動をフィードバックするために有用な実時間調理行動認識手法を提案した．提案手法では，両手首の加速度センサから得た加速度データを，畳み込み LSTM により 1 秒間に 1 回の頻度で実時間推論を実行する．モデルを学習するためのデータは，単一の被験者が加速度センサを装着した状態で実際にカレーを調理して取得したデータより作成した．実際にカレーを調理しながら実時間推論した結果と，テストデータをバッチ推論した結果とを比較した．

今後は，学習データの作成方法を改善することにより，実時間推論の性能を向上させることができると考える．また，被験者の数を増やして，より汎用的に使用できるように改善すること，調理行動のクラス数を増やして適用できる料理のレパートリーを拡大することも必要である．

参考文献

1. Shuichi Urabe MY Katsufumi Inoue. Cooking activities recognition in egocentric videos using combining 2DCNN and 3DCNN. CEA/MADiMa ’18: Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management. 2018; 1–8.
2. Sho Ooi MS Tsuyoshi Ikegaya. Cooking behavior recognition using egocentric vision for cooking navigation. Journal of Rebotics and Mechatronics. 2017;29: 728–736.
3. Shinya Michibata MY Katsufumi Inoue. Cooking Activity Recognition in Egocentric Videos with a Hand Mask Image Branch in the Multi-stream CNN. CEA ’20: Proceedings of the 12th Workshop on Multimedia for Cooking and Eating Activities. 2020; 1–6.
4. Yusaku Korematsu NM Daisuke Saito. Cooking State Recognition based on Acoustic Event Detection. CEA ’19: Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities. 2019; 41–44.
5. Yanli Ji AS Yoshiyasu Ko. Cooking Gesture Recognition using Local Feature and Depth Image. CEA ’12: Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities. 2012; 37–42.
6. Yuma Hijioka TK Makoto Murakami. Motion analysis for cooking motion recognition. 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). 2016; 551–555.
7. T.Kato AY Hanano Kobayashi. Proposing a system to Supporting for Kitchen Knife Skill Improvement using Acceleration Sensor. 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). 2020; 79–80.
8. 加藤 岳大 et al. SynCook：動画メタデータと加速度センサを用いたレシピ動画進行度自動同期システム. 日本ソフトウェア科学会 WISS 2020. 2020; 5 pages.
9. Sayeda Shamma Alia ST Paula Lago. Summary of the cooking activity recognition challenge，human activity recognition challenge. Siringer. 2020.
10. Isibor Kennedy Ihianle SHE Augustine O. Nwajana. A deep learning approach for human activities recognition from multimodal sensing devices. IEEE Access. 2020;8: 179028–179038.

Open Access This article is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/