

令和 5 年 6 月 19 日現在

機関番号：12601

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K11369

研究課題名（和文）音環境理解と対話管理技術に基づく大規模レシピデータを用いた調理支援に関する研究

研究課題名（英文）Research on cooking support using large-scale recipe data based on sound environment understanding and dialogue management technology

研究代表者

齋藤 大輔（Saito, Daisuke）

東京大学・大学院工学系研究科（工学部）・准教授

研究者番号：40615150

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究では調理行動において発生する音の情報に着眼して、音響イベント検出技術の改良と応用に取り組みことで、調理動作を妨げない適切な情報提示としての調理支援システムの実現を検討した。特に実環境における調理音の取り扱いについて、収録から単一の調理行動推定、さらには調理全体を通じた時系列の調理行動推定の実験を行うことで、調理音に対する音響イベント検出技術がこれらの支援に活かせる可能性を示した。またレシピのテキストデータについて、それを調理支援の対話シナリオに変換する新しい技術について、その初期となるフローグラフ変換と動詞情報のクラスタリングを行い、その有効性を示した。

研究成果の学術的意義や社会的意義

本研究は調理という人間の普通の生活に密に関わるアクティビティに対して、音声情報処理技術がどのように貢献できるかについて示したものであり、音声情報処理技術の学術的拡張という側面、および社会全体における人間の生活の充実という社会的貢献の意味とともに大きな意義を持つものであるといえる。今後は近年急速に発展するAIの基盤技術と本研究課題における知見をいかに融合していくかが課題であり、それを通じて更なる学術的意義の拡張、社会的貢献を実現するといえる。

研究成果の概要（英文）：In this study, we focused on the sound information generated during cooking behavior, and studied the improvement and application of acoustic event detection technology to realize a cooking support system that presents appropriate information without interfering with cooking behavior. In particular, we conducted experiments on the handling of cooking sounds in a real environment, from recording to estimation of a single cooking behavior, and then to estimation of a time series of cooking behaviors throughout the entire cooking process, and demonstrated the possibility that acoustic event detection technology for cooking sounds can be used to support these activities. We also demonstrated the effectiveness of a new technique for converting textual recipe data into dialog scenarios for cooking support by performing the initial flow graph transformation and clustering of verb information

Translated with www.DeepL.com/Translator (free version)

研究分野：音声言語情報処理

キーワード：調理行動支援 音響イベント検出 行動推定

1. 研究開始当初の背景

インターネットにおいて、ブログや SNS などの、ユーザ主導の情報発信がごく一般的なものとなる中で、料理レシピを公開したり、それを共有することが広く行われるようになってきている。例えば 1998 年にサービスを開始したレシピコミュニティサイトであるクックパッド (<http://cookpad.com>) には、2017 年 10 月現在、240 万以上のレシピが登録されており、月間で 6000 万人の利用者数がある。このような背景から、近年ではスマートフォンなどを片手にこれらのレシピを確認しながら調理するなどといったことも広く行われるようになった。しかし通常、調理とスマートフォン等の操作を両立して行うのは困難を伴う。加えて料理には適切な時間やタイミングが存在することから、調理動作を妨げない適切な情報提示が必要となる。またレシピが、大規模な「情報」として存在する現状においては、情報処理技術によって、これらの情報をユーザが利用しやすい形式で提供し、ユーザの調理体験の質を向上させることが望まれている。

2. 研究の目的

食に対する情報処理としては、食画像の画像認識や画像からの摂取カロリーの推定など、食に関する「静的な情報」に対する情報処理が、主に画像処理の分野において研究されている。一方で、レシピの「情報」とユーザの調理体験そのものを適切に結びつけるための技術はまだあまり研究されていない。これを実現するためには、レシピ情報を「動的なシナリオ」として解釈する必要があり、このシナリオとユーザの調理体験を対応付けるようにシステムが支援するといった形式が必要となる。すなわち、どのようにして膨大な静的記録情報と実世界の動的活動とを融合・混合し価値を創成するかという学術的な「問い」に対して答える必要がある。本研究では音声対話システムで培われた「音声対話に基づくユーザへの情報提示」の技術を、調理支援に応用することで、この問いへの解答を与える。

本研究は、音声言語情報処理ならびに音声インタラクションの研究において培われてきた音声対話システムのための基礎技術、及び日常にあふれる様々な音のイベントを検出・認識する音環境理解の技術を基盤として、大規模な料理のレシピデータとユーザの調理行動を結びつけることで、ユーザの調理活動を支援・活性化するシステム、ならびにこれまで静的なものであったレシピデータをユーザの調理行動を介して動的に最適化する技術の研究及び開発を目的とする。

3. 研究の方法

(1) 実環境での調理音による調理行動推定

これまでに公開されている調理行動のデータベースは主に映像や画像処理を意図しており、音響データを併用するような場合でも十分な知見が得られてこなかった。そこで決められたメニューに対して、一連の調理行動を行った際の音響データの収録を行い、この整備を行う。加えて収録された音響データを用いて音響イベント検出技術を組み合わせた調理行動推定を行う。

(2) レシピに沿った調理支援のためのレシピテキストの構造化

レシピデータを対話シナリオに変換する技術はレシピに沿った調理支援には必要不可欠な技術である。複数のレシピに登場する調理行動に相当する動詞を抽出し整理することで、効率的な変換が実現しうる。さらにレシピのテキストデータをフローグラフと呼ばれる形式に変換することで抽象的な演算を通して複数レシピを統合することなどが可能になる。

(3) 隠れマルコフモデルを用いた時系列イベント検出による調理行動推定

レシピに基づいて時間的な音響イベントの変化を予測することは調理行動全体の推定には有用と考えられる。本研究では隠れマルコフモデルを用いた時系列認識モデルを採用しその効果を検証していく。

(4) 調理行動認識における音響識別器の中間特徴量の利用

ニューラルネットワークによる音響識別器の中間特徴量の利用という新しいアプローチを検討する。これまでの調理行動推定の枠組みでは調理行動という限定された行動におけるデータを用いて学習が行われていた。ここに汎用的な識別器の性能とデータ量を生かすため、汎用的な音響識別器を特徴量抽出器の形で利用することで、特定ドメインである調理行動認識・調理音響イベント検出においてその精度を向上させることが可能となる。

4. 研究成果

(1) 研究の主な成果

① 実環境での調理音による調理行動推定

調理音を「切る」、「炒める」、「その他」の3つのクラスに分類する識別実験を4つの手法で行った。録音した6つの収録データを用いて6-fold cross validationを行なった。全ての実験について確率密度関数は混合ガウス分布 (Gaussian Mixture Model; GMM) を用いた。

まず、ベースラインとして、調理音を特徴量系列に変換し、各時間フレームが常にGMMに従うと仮定して識別実験を行った(MFCC)。特徴量は12次元MFCCに1次と2次の動的特徴量を結合した36次元のベクトルで、フレーム長・フレームシフトはどちらも16ms、評価時には長さが1秒になるようにフレームをセグメントにまとめ、セグメントごとに識別を行った。次に、全ての調理音データについて、GMMによるソフトクラスタリングを行い、フレームごとのクラス事後確率を用いてヒストグラムを生成する手法を試みた(GMM)。GMMの混合数は20で、ヒストグラムは1秒ごとに生成されるように設定した。3つ目として、提案手法であるNMF基底ヒストグラムを用いた(NMF)。NMFの規定数は20で、ヒストグラムは1秒ごとに生成されるように設定した。最後に、提案手法に加えて、NMFで基底分解する前にスペクトログラムを二つのスペクトログラムに分解する手法も検討した(HPSS)。分解手法は、メディアンフィルタによって時間方向と周波数方向の連続性をそれぞれ保存することを目的とするもので、これによりスペクトル基底が適切に学習されることが期待される。NMFの基底数は分解した二つのスペクトログラムそれぞれ10個で、ヒストグラムは1秒ごとに生成されるように設定した。GMMの混合数はMFCCのみ2混合で、他の3つの手法は32混合とした。

結果としてNMFを特徴量として用いることで62%程度の精度で調理行動の識別が可能となった。

② 調理行動認識における音響識別器の中間特徴量の利用

EnvNet-v2と呼ばれる音響識別器にボトルネック層を挿入し特徴量抽出器としたEnvNet-v2をESC-50と呼ばれる汎用的な音響識別タスクのデータで学習させて、得られたネットワークをもとに特徴量抽出器を構成した。結果として炒めるなどの調理行動に対しては5ポイントほどの識別率の改善が見られ、特に時系列情報を考慮した場合には90%ほどの精度で調理行動を識別することが可能になった。

(2) 得られた成果の位置づけ

本研究は調理支援というタスクにおいて、音声情報処理技術がどのように貢献できるかについてその可能性を示したものであるといえる。

(3) 今後の展望

近年のニューラルネットワーク技術の進展より、本研究課題の実施前よりはるかに技術融合が盛んになってきている。今後は人間の行動の推定および支援というより広い枠組みで本研究の成果を出発点に最新の技術との融合を図っていくことが望まれる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 宮澤要二, 齋藤大輔, 峯松信明
2. 発表標題 レシピ情報に基づく調理行動認識における音響識別器の中間特徴量利用に関する検討
3. 学会等名 音学シンポジウム2020
4. 発表年 2020年

1. 発表者名 Yusaku Korematsu, Daisuke Saito, Nobuaki Minematsu
2. 発表標題 Cooking State Recognition based on Acoustic Event Detection
3. 学会等名 CEA'19 (国際学会)
4. 発表年 2019年

1. 発表者名 是松優作, 齋藤大輔, 峯松信明
2. 発表標題 調理支援のための実環境での調理音による調理行動推定
3. 学会等名 日本音響学会秋季研究発表会
4. 発表年 2018年

1. 発表者名 是松優作, 齋藤大輔, 峯松信明
2. 発表標題 レシピに沿った調理支援のためのレシピテキストの構造化に関する基礎検討
3. 学会等名 IDR (情報学研究データレポジトリ) ユーザフォーラム2018
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	山岸 順一 (Yamagishi Junichi) (70709352)	国立情報学研究所・コンテンツ科学研究系・教授 (62615)	
研究 分担者	高木 信二 (Takaki Shinji) (50735090)	国立情報学研究所・コンテンツ科学研究系・特任助教 (62615)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------