

# 前腕動作に着目した食事内容推定技術 — 手軽な食事管理をめざして —

先進技術研究所

さいとう  
齊藤たかど  
隆仁かわさき  
川崎さとし†  
仁嗣いけだ  
池田だいぞう  
大造かたぎり  
片桐まさじ  
雅二

近年、過食や偏食による生活習慣病が社会問題として顕在化している。食事バランスを管理するサービスは多数あるものの、ユーザ自身により食事に関する情報を欠損なく継続的に記録する必要があり、手間がかかるため、サービスから適切な支援を受けるための障壁となっていた。本研究では、摂食に伴う前腕動作に着目した食事内容推定手法を提案する。利き手に着用したウェアラブルデバイスのセンサデータから食事内容を推定し、ユーザ入力の手間をかけずに食事内容を把握することが可能となる。

## 1. まえがき

近年、偏食や過食などの食生活の乱れが生活習慣病を引き起こす一因であると言われ、社会問題として顕在化している。ライフスタイルの多様化や、食事の選択肢の拡大に伴い外食のメニューや加工食品を摂取する機会が増加しているが、これらの食品の大半は食塩や脂肪を多く含む傾向があり、摂取し続けることで生活習慣病の予備軍が増加すると懸念されている。外食や加工食品を日常生活に取り入れる際には、食事バランスを考慮することが重要である。

農林水産省は国民の健康増進を目的として、食生活指針 [1] を策定し、食生活を習慣的に見直すこ

とを提唱している。また厚生労働省は、食生活指針を具体的な行動に結びつけるために、食事バランスガイド [2] を策定している。このような取組みがある一方、内閣府の調査によると、2人に1人は適切な食生活を送ることができておらず [3]、その理由としては、①何をどのように実行したらよいか分からない、②忙しくて時間が無い、といったことが上位にあげられている [4]。

このような状況の中、食事バランスの管理を支援するサービスが多く企業の企業や団体から提供されている。サービスの提供形態は多岐にわたり、スマートフォンを利用した食生活支援サービスも多数提供されている [5]。これらサービスを利用してユーザが

©2017 NTT DOCOMO, INC.

本誌掲載記事の無断転載を禁じます。

† 現在、サービスインベション部

適切な支援を受けるためには、日々の食事に関する情報を、欠損なく継続的に記録することが望ましい。しかしながら既存サービスの多くは、食事を取った時間や、料理品目を示す食事内容の入力をユーザ自身が手動で行う必要があり、その手間から、継続的な利用の障壁になっている。

食事を取った時間に関しては、手首の動作から高精度に認識する技術 [6] が開発されており、このような機能を有するウェアラブルデバイス [7] が実用化されている。一方、食事内容の記録に関しては、カメラで撮影した食事画像を用いて認識する手法 [8] や、摂食\*1に伴う上肢動作を用いて認識する手法 [9] などが研究されている。前者は、撮忘れや利用者のプライバシー保護が課題となっており、後者は複数箇所にデバイスを着ける必要があるため日常的に利用する上で大きな阻害要因となっている。

そこでドコモは、手間をかけずに食事内容を継続的かつ自動で把握することを目的とし、上肢動作のうち前腕動作のみで食事内容を推定する技術を開発した。本稿では、リストバンド型の汎用的なウェアラブルデバイス1台を利き手の前腕に着用し、取得されるセンサデータから食事内容の認識を実現する手法について解説する。

## 2. 関連研究

### 2.1 Amftらの手法

食事内容を推定する手法として、Amftらは、両腕の上腕および前腕の合計4カ所に加速度センサ\*2および角速度センサ\*3を装着し、これらセンサから取得した上肢動作のデータと、スロートマイク\*4およびイヤホン型マイク\*5から取得したデータを利用した手法を提案している [9] [10]。彼らの手法は、食べ物を切るような準備動作や、口に運ぶ動作、咀嚼動作、嚥下動作といった食事で発生する一連動作

のサイクルを各センサにより検出することで食事内容を推定し、画像以外のデータを用いることの有効性を示した。この研究では、食事内容を推定するためには、時系列的な動作のサイクルが重要であることが示唆されている。一方、この手法では、複数箇所にウェアラブルデバイスを用いて高い精度で状態を検出し、状態遷移モデル\*6を利用した手法で食事内容を推定しているが、複数デバイスの装着が必要となることから、日常的に利用する上で大きな阻害要因となってしまう。

### 2.2 Zhangらの手法

Zhangらの歩行や走行など人物の動作を認識する技術の研究では、状態の遷移ではなく状態の出現頻度に着目し、人物動作を高精度で認識する手法を提案している [11]。この手法では、歩行や走行などの動作は、基本要素となる特徴的な動作の複数集合であると解釈し、BoW (Bag-of-Words) 表現\*7を適用している。この関連研究が示すように、食事における前腕動作を認識する際にも、人物動作に着目するという観点で見ると、BoW表現が有効な可能性があると考えられる。そこで、本研究ではBoW表現を用いて、前腕動作を基本要素となる特徴的な動作の組合せとして認識する方法を用いる。

## 3. 提案手法

### 3.1 概要

提案手法の概要を図1に示す。本研究は、摂食時の前腕動作が食事内容によって異なることに着目し、前腕動作の違いから食事内容を推定する。例えば、ハンバーガーの摂食時は、口に運ぶ動作や噛みしめる箇所を変える動作が生じる。一方で、ラーメンの摂食時は、麺を持ち上げる動作や、麺を冷ます動作、吸る動作などが時系列的に生じる。これらの食事内

\*1 摂食：食べ物を口まで運び食べること。

\*2 加速度センサ：速度の変化を計測するセンサのこと。携帯端末に搭載することで、携帯端末の動きの変化を検出することが可能となる。また、センサに対して固定された、互いに直交する3方向の加速度を測定できるセンサを3軸加速度センサとよぶ。

\*3 角速度センサ：回転方向の速度を計測するセンサのこと。

\*4 スロートマイク：喉元に着用するマイクで、小さな音声や、嚥下音などが集音できるマイクのこと。

\*5 イヤホン型マイク：咀嚼に伴って発生する肉伝導音を集音することができるマイクのこと。

\*6 状態遷移モデル：ある手続きの流れを、開始状態から始まり、いくつかの状態を遷移して終了状態に辿り着くように表現したモデルのこと。ここでは食事という手続きは、準備動作を開始状態として、摂食、咀嚼の状態を繰返し遷移し、嚥下といった終了状態に辿り着く。

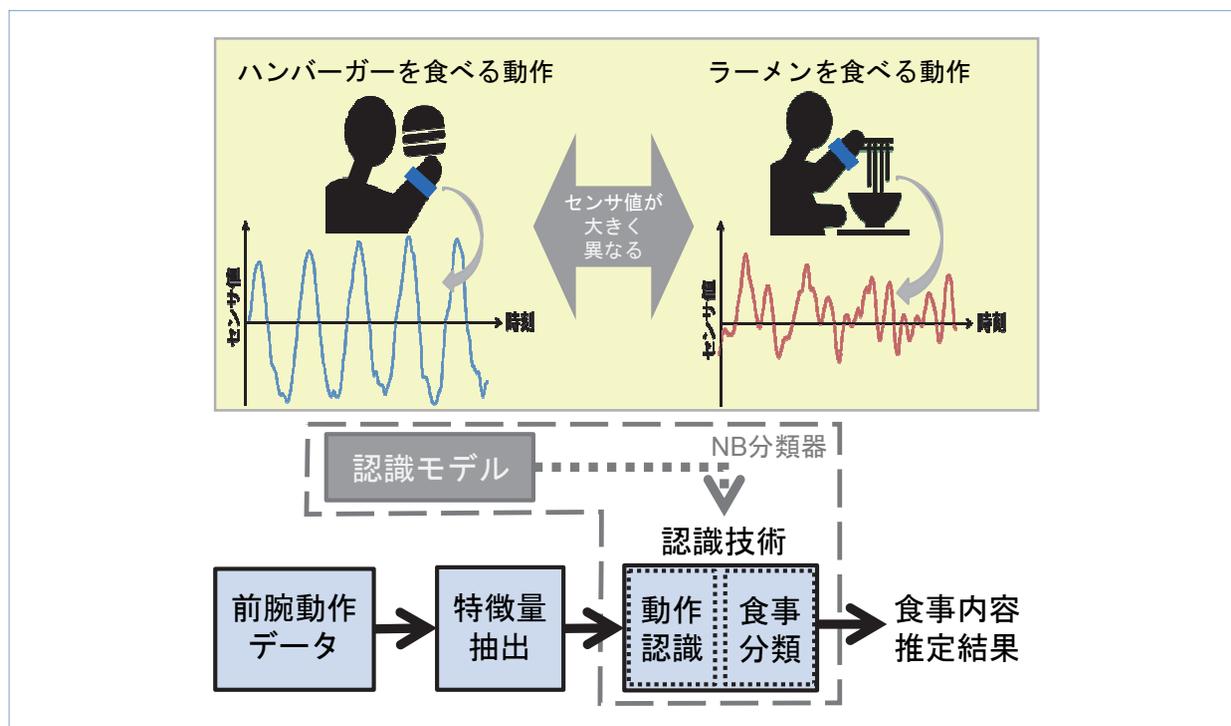


図1 提案手法の概要図

容による動作の違いを、動作が発生する順序を時系列情報として考慮し認識することで食事内容を推定する。提案手法では、利き手の前腕にウェアラブルデバイスを着用し、摂食時の前腕動作に伴うセンサデータ（以下、前腕動作データ）を取得する。この前腕動作データから特徴量<sup>\*8</sup>を抽出し、Naïve Bayes分類器<sup>\*9</sup>（以下、NB分類器）によって動作認識を行い、食事内容を推定する。

前述の従来手法では複数のウェアラブルデバイスの着用が必須であり、1台で同様の手法を実現することは困難であった。そこで、提案手法では、摂食時の動作を、基本要素となる特徴的な動作の集合として捉えるBoW表現を用いる。そして動作の発生順序によって時系列情報を考慮できるN-gram<sup>\*10</sup>に拡張したアプローチによって、従来手法では難しかったウェアラブルデバイスの数を低減させた食事

内容推定を実現する。

本手法により、市販のリストバンド型デバイス1台を利き手に装着するだけで、ユーザに手間をかけさせず食事内容を継続的かつ自動で把握する技術を実現した。

### 3.2 推定の流れ

3軸加速度センサのデータを前腕動作データとして使用し、これにフィルタ処理を施しデータのノイズを低減させた後、一定幅の移動窓<sup>\*11</sup>内のセンサデータより特徴量を抽出する。特徴量の次元数を不必要に増加させると分類器の性能低下に繋がるため、本研究では、表1に示す加速度センサの統計量の12次元を特徴量として用いる。摂食時の前腕動作では、動作が発生する順番を考慮した認識手法が有効であると考えられる。そこで、まずこの12次元の特徴量

\*7 BoW表現：文書のカテゴリ分類に利用される手法であり、文書に登場する各単語の出現頻度によって、その文書を表現する方法のこと。

\*8 特徴量：データを比較するために、任意の方法で算出した数値やその集まりのこと。また、元となるデータから特徴量を算出することを特徴量抽出という。3軸加速度センサのデータのように、データが複数列で構成される場合や、算出する方法を複数選んだ場合は、特徴量も複数列となることがある。このと

き、特徴量を構成する列数を次元数という。

\*9 Naïve Bayes分類器：ベイズ確率と呼ばれる確率論に基づいた分類器のこと。

\*10 N-gram：文書に含まれる単語のうち、連続するN個を新たな単語とみなす方法をいう。

\*11 移動窓：センサデータに処理を施すとき、対象とするデータの範囲を窓と呼び、その窓を一定のルールに従って移動させたものを移動窓という。

に対してBoW表現を適用し、基本要素となる特徴的な動作（以下、要素動作と呼ぶ）の集合として表現する（図2①）。その上で、要素動作が発生する順番を考慮して、複数の要素動作をまとめたもの（以下、一連要素動作と呼ぶ）の集合を考え（図2②）、前腕動作データを一連要素動作が発生する頻度で表現し（図2③）、NB分類器に適用することで食事内容を推定する。図2に示すように、前腕動作

データを要素動作の集合で表現したのち、N-gramに拡張することで、一連要素動作の集合を表すことができる。このとき要素動作は“words”として表現され、一連要素動作としては、例えば、麺を啜る動作や丼ものを口に掻き込む動作が表現されることを想定している。麺を啜る動作を例にとると、麺をつかむ動作、上に持ち上げる動作、口に運ぶ動作などによって構成される。この各動作を時間軸に沿ってさらに細分化したものが要素動作として表現されることになる。なお、口に運ぶ動作や咀嚼中に前腕を動かさない動作はあらゆる食事内容で発生するため、このようなwordsに着目しても、食事内容の推定に対しては有効となり難い。そこで、あらゆる食事内容の摂食動作で共通して頻出するwordsは、その重要度を軽減させ、食事内容固有の摂食動作として発生しやすいwordsはその重要度を大きくすることで、

表1 利用する特徴量

特徴量	次元数
平均値	3
標準偏差	3
分散	3
中央値	3

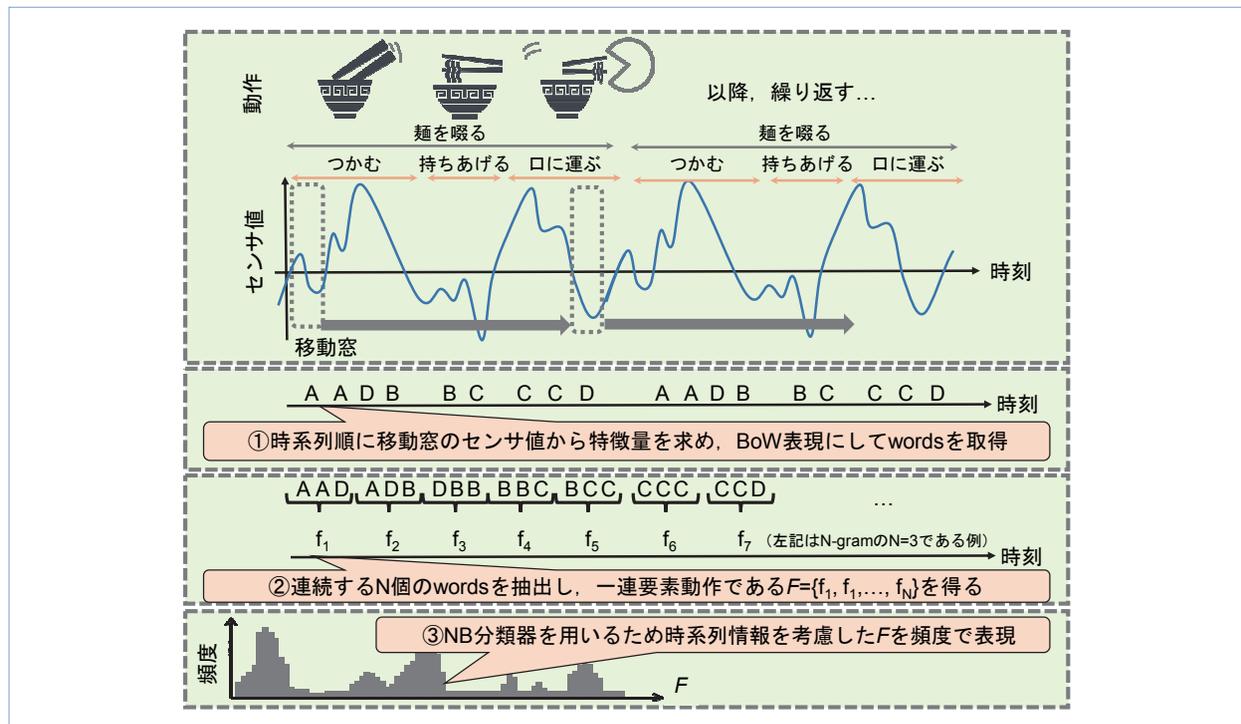


図2 N-gramを用いたBoW表現

食事内容の推定精度を高める。

摂食時には要素動作が繰り返し発生したり、交互に行われたりするが、前述のAmftらの手法は1回の食事に含まれる状態遷移すべての連続性を考慮することになる。一方で、NB分類器を用いた動作の頻度に着目した本手法は局所的な連続性を捉えることが可能である。そのため、突発的に生じるノイズや食事内容で食べる順序が変わる場合でも、従来手法に比べて食事内容の推定精度への影響を抑制する効果が期待できる。

## 4. 実験および精度評価

### 4.1 前腕動作データの収集および実験条件

#### (1) 実験条件

認識対象とする食事内容は、表2に示す5品目とし、また被験者が摂食時に利用する食器具および収集したデータ数を表に記載した。Android Wear™\*12を搭載するリストバンド型デバイス（HUAWEI WATCH W1®\*13）によって前腕動作データを取得する。被験者は図3に示すように、食器具を利き手

表2 食事内容と食器具およびデータ数内訳

食事内容	食器具	要求条件
丼もの（牛丼、かつ丼、親子丼、天丼、海鮮丼）	箸	29
カレーライス	スプーン	32
パン類（菓子パン、サンドウィッチ、ピザ、ハンバーガ）	素手	26
パスタ（スパゲッティ）	フォーク	27
麺（ラーメン）	箸	33

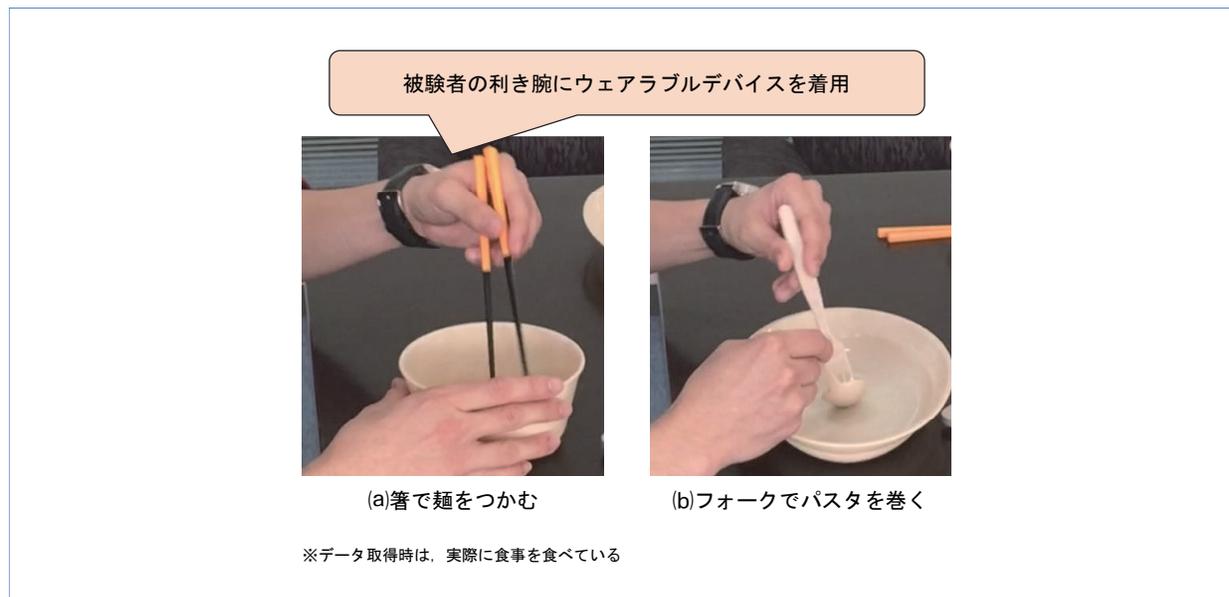


図3 前腕動作データの取得実験の様子

\*12 Android Wear™ : Google Inc. の商標.

\*13 HUAWEI WATCH W1® : HUAWEI Technologies Co. の登録商標.

に把持した状態で食事をとり、その際の前腕動作データを収集した。摂食時の前腕動作データは1食当たり約7分間のデータである。被験者は20～40代の計9人（男性7人、女性2人）であり、男性のうち1名が左利きで他の被験者は右利きである。なお、Amftらの研究 [9] で用いられたデータ数を参考に、食事内容1品目に対して約30食分をめざして前腕動作データの収集を行った。ただし、被験者の日常生活で指定の食事内容の前腕動作データを収集しているため、被験者や食事内容によってデータ数に差がある。また、この前腕動作データは職場の食堂や自宅、レストランなどで取得しており、同一の食事内容でも全く同じものを摂食したデータとは限らない。

#### (2) 実験データの収集

一般に身体が行う動作は99%が15Hz未満であると言われている [12]。そのため、加速度センサを用いて、歩行やランニングなどの動作を認識する場合は余裕をみて100Hz程度の周波数で計測することが多い [11]。また、実生活で手首に着用したウェアラブルデバイスから得られる食事動作の周波数は0.2～0.6Hzであることが示されている [13]。サンプリングするときの周波数は対象とする動作の周波数の2倍以上であることが望ましいため、本研究では食事動作の周波数の最大値である0.6Hzから十分余裕をもたせた20.0Hzで取得した前腕動作データを用いて、特徴量を抽出する。

#### (3) BoW表現の適用

抽出した特徴量から要素動作（words）を収録したcodebook\*14を作る。codebookが含む全wordsの数をVocabularyと呼ぶことにする。動作認識にBoW表現を適用する場合は、比較的Vocabularyを小さい値にすることが効果的であると示されている [11]。そこで前腕動作データから抽出した特徴量に対してVocabulary = 20としたBoW表現を適用する。

#### (4) N-gramの適用

BoW表現とした後、発生したN個のwordsに注目しN-gramを適用することで、麺を吸る動作やパスタをフォークで巻く動作のように、時間幅をもつことで意味を成す動作の表現を試みる。N-gramを適用するとwordsの組合せが爆発的に増加するため、本研究ではN = 3のTrigram\*15を採用する。この値は、麺を吸る動作やパスタをフォークに巻く動作などの、継続する時間幅を考慮して設定されたもので、20.0Hzでサブサンプリング\*16した前腕動作データに対して1秒分の移動窓を0.5秒ずつ移動させて2秒分のサンプルを表現することになる。

## 4.2 精度評価

本実験では、収集した前腕動作データをNB分類器に学習させ、その推定性能を測る。推定性能の評価方法は、LOOCV (Leave-One-Out Cross Validation)\*17とした。この評価方法は、収集したすべての前腕動作データの中から1食分のデータを評価用とし、残りのデータを分類器の学習用とする方法である。評価指標は一般的なRecall（再現率\*18）、Precision（適合率\*19）の調和平均であるf1値\*20および、前腕動作データの総数に対する食事内容ごとの真陽性\*21と真陰性\*22の総数の比である正解率Accuracyを用いる。

LOOCVの混同行列を図4に示す。この表は、縦方向に前腕動作データが属する食事内容の真値を示し、横方向に分類器の推定結果を示すものである。背景色が黒色に近いほど推定結果の中で占める割合が高い。混同行列から、特に麺類の推定結果が他の食事内容に比べて高いことが分かる。一方で、カレーライスはパスタや麺として誤って推定されることが多かった。これは、本実験では手首の回転を直接測定できる角速度センサを用いなかったため、手首の回転に関する特徴が反映され難かったことが原

\*14 codebook：BoW表現で利用したすべてのwordsを集めた索引であり、dictionaryとも呼ばれる。

\*15 Trigram：N-gramにおいてN = 3としたときの名称であり、N = 2のときをBigram、N = 1のときをUnigramと呼ぶ。

\*16 サブサンプリング：センサデータから一定の規則に沿って、その一部を抜き出すこと。

\*17 LOOCV：分類器の評価に用いるデータを分割する方法のひとつ。全データを余すことなく評価用として利用するため、データ総量が少ない場合に利用されることが多い方法である。

\*18 再現率：推定結果の漏れの少なさを表現することができ、網羅性を意味する。しかし、推定結果の正確さを表現することができない。

\*19 適合率：推定結果の正確さを表現することができる指標のこと。しかし、推定結果の網羅性を表現することができない。

困として考えられる。具体的には、カレーライスを食べる際には、カレーのルーとライスを寄せ集める動作の後、口に運ぶ動作が発生する。このとき、表1で採用した特徴量では、カレーのルーとライスを寄せ集める動作が口に運ぶ動作と同一視され、他の食事内容との区別が難しくなった可能性があり、事

前確率<sup>\*23</sup>が最も高い麺として誤推定することが多くなったと考えられる。

食事内容推定の結果として、図5にf1値、図6にAccuracyを示す。食事内容に対するf1値の平均値は63%で、Accuracyの平均値は72%となった。カレーライスはRecall、Precisionともに著しく低いた

		推定結果				
		丼もの	カレー	パン類	パスタ	麺
真 値	丼もの	0.55	0.00	0.00	0.17	0.28
	カレー	0.16	0.09	0.00	0.34	0.41
	パン類	0.04	0.04	0.68	0.12	0.16
	パスタ	0.19	0.04	0.04	0.62	0.08
	麺	0.03	0.00	0.00	0.06	0.91

図4 食事内容推定：混同行列

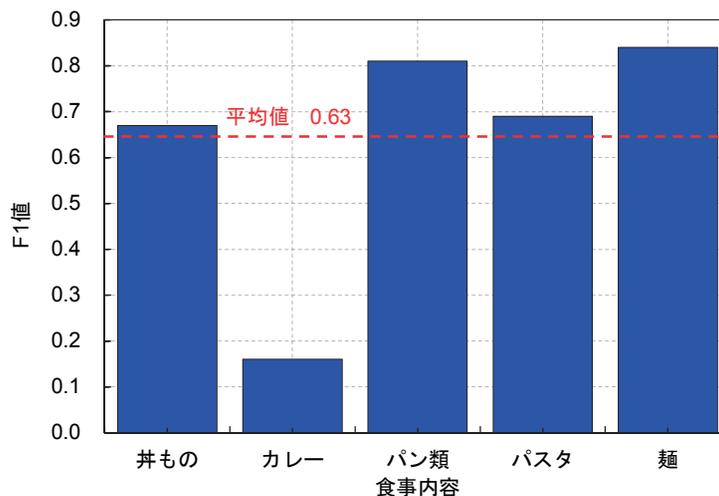


図5 食事内容推定の結果 (f1値)

\*20 f1値：再現率と適合率の両指標を調和平均によって統合した指標のこと。再現率と適合率共に高い分類器が最良であるが、両指標はトレードオフの関係がある。そのため個々の指標で評価するのではなく総合的な評価が可能であるf1値を用いることが多い。

\*21 真陽性：真値と推定結果が一致する結果のうち、実際にとった行動と推定結果が一致しているもの。例えば、実際にラーメンを食べたときに、その推定結果もラーメンである事象の数。

\*22 真陰性：真値と推定結果が一致する結果のうち、実際にとっていない行動と推定結果が一致しているもの。例えば、食事内容がラーメンでないときに、その推定結果もラーメンでないことと推定する事象の数。

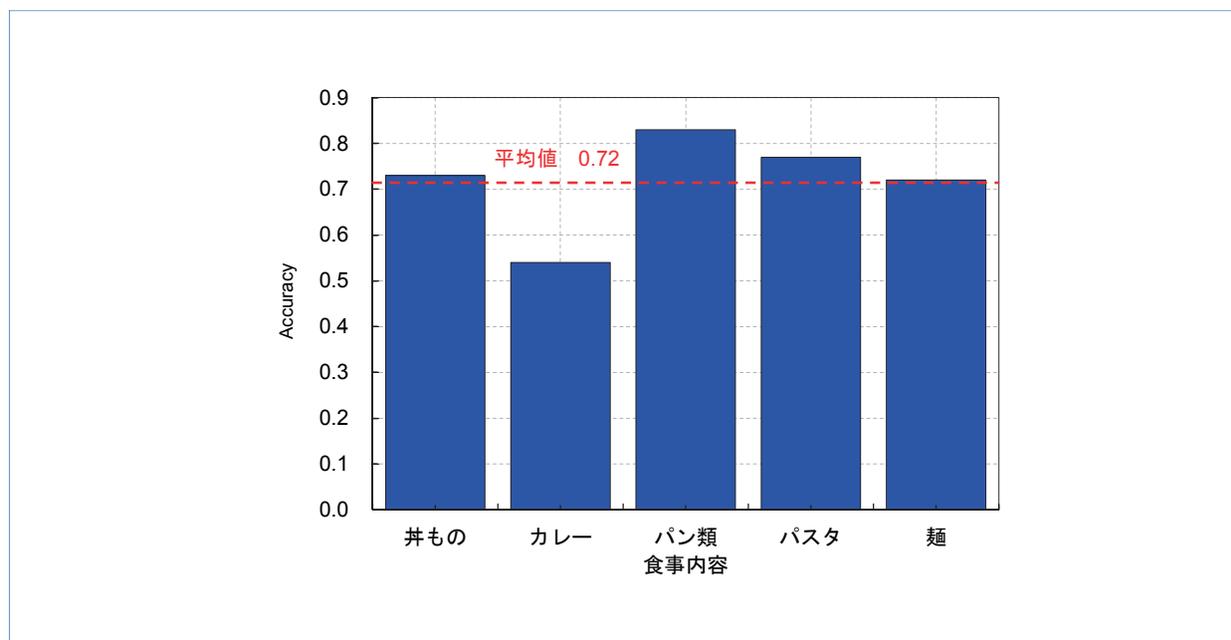


図6 食事内容推定の結果 (Accuracy)

め、f1値およびAccuracyの両方で最も推定精度が低い。その他の食事内容は、f1値は65%以上、Accuracyは70%以上を保っており、食事の品目がある程度限定できる状況では、前腕動作データによって食事内容推定が可能であることが示された。

## 5. あとがき

ユーザに手間をかけさせずに食事内容を継続的かつ自動で把握する技術の実現をめざし、前腕動作に着目した食事内容の推定手法について検討した。提案手法の認識精度は本実験条件において、f1値で平均63%、Accuracyは平均72%を達成した。これは市販のリストバンド型ウェアラブルデバイス1台で、職場の食堂のような食事の品目がある程度限定できる状況では食事内容推定が可能であることを示している。今後は、本技術を適用したサービスの実用化に向け、外食のメニューや自宅における食事のよう

に広範な品目まで推定対象を拡大させるとともに、認識精度向上をめざす。本研究では、特徴量として一般的な統計量を採用したが、一部の食事内容の推定精度が低いことから、前腕動作をより詳細に表現できる特徴量の追加が必要といえよう。また、同一の食事内容においても、被験者間の前腕動作に違いがあることを確認しており、今後の精度向上には個人ごとに認識モデルや特徴量の調整を行うことが有効であると考えられる。

### 文献

- [1] 農林水産省：“食生活指針について。”  
<http://www.maff.go.jp/j/syokuiku/shishinn.html>
- [2] 厚生労働省：“「食事バランスガイド」について。”  
<http://www.mhlw.go.jp/bunya/kenkou/eiyou-syokuji.html>
- [3] 内閣府：“食育に関する意識調査報告書（平成28年3月）。”  
<http://www.maff.go.jp/j/syokuiku/ishiki/h28/>
- [4] 厚生労働省：“平成26年版厚生労働白書 健康長寿社会の実現に向けて～健康・予防元年～。”

\*23 事前確率：ベイズ確率では、ある事象が発生する確率は、その事象に関する知識量によって変化するという考えに基づいた確率論である。知識を得る前に、ある事象が発生すると考えていた確率を事前確率といい、知識を得た後の確率を事後確率という。

- <http://www.mhlw.go.jp/wp/hakusyo/kousei/14/>
- [5] NTTドコモ：“dヘルステアバック.”  
<https://www.nttdocomo.co.jp/service/dmarket/healthcare/>
- [6] Y. Dong, J. Scisco, M. Wilson, E. Muth and A. Hoover: “Detecting Periods of Eating During Free-Living by Tracking Wrist Motion,” Proc. of IEEE Journal of Biomedical and Health Informatics, Vol.18, No.4, pp.1253-1260, Sep. 2013.
- [7] TDK株式会社：“SilmeeW20.”  
[https://product.tdk.com/info/ja/products/biosensor/biosensor/silmee\\_w20/index.html](https://product.tdk.com/info/ja/products/biosensor/biosensor/silmee_w20/index.html)
- [8] K. Aizawa, Y. Maruyama, L. He and C. Morikawa: “Food Balance Estimation by Using Personal Dietary Tendencies in a Multimedia Food Log,” Proc. of IEEE Transaction on Multimedia, Vol.15, No.8, pp.2176-2185, Dec. 2013.
- [9] O. Amft, M. Kusserow and G. Tröster: “Probabilistic parsing of dietary activity events,” Proc. of International Workshop on Wearable and Implantable Body Sensor Networks, Vol.13, pp.242-247, 2007.
- [10] O. Amft and G. Tröster: “On-Body Sensing Solutions for Automatic Dietary Monitoring,” Proc. of IEEE Pervasive Computing, Vol.8, No.2, pp.62-70, 2009.
- [11] M. Zhang and A. A. Sawchuk: “Motion Primitive-Based Human Activity Recognition Using a Bag-of-Features Approach,” Proc. of the 2nd ACM SIGHIT International Health Informatics Symposium, pp.631-640, Jan. 2012.
- [12] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell and B. G. Celler: “Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring,” Proc. IEEE Trans. on Information Technology in Biomedicine, Vol.10, No.1, pp.156-167, Jan. 2006.
- [13] 矢野 和男, 栗山裕之：“「人間×センサ」センサ情報が変える人・組織・社会,” 日立評論, Vol.89, No.07, pp.62-67, Jul. 2007.