4章 食メディアに関する研究最先端

正会員 井手 一郎^{†1}, 柳井 啓 司^{†2}, 山 肩 洋 子^{†3}

キーワード：食メディア，料理画像，料理レシピ

1. まえがき

一昔前，台所は家庭内の情報化における「最後の聖域」と言われていた．その後も，家電各社からインターネット冷蔵庫やインターネット電子レンジなどの製品が時折現れてはいつの間にか消えているという状況で，今日でも情報化が進んだとは言いがたい．強いて言えば，スマートフォンやタブレットなどの携帯情報端末を持ち込んで，料理レシピを閲覧するユーザが増えたくらいではなかろうか．

一方，食に関するコンテンツ（以下，食メディア）を対象とした情報処理技術はこの10年間で大きく発展した．これは，携帯情報端末の普及と，マイクロブログや料理レシピ投稿サイトなどウェブ上のソーシャルメディアサービスの普及に伴い，食事画像や料理レシピなど，研究データとして扱うことができる食メディアの量が増えたこと，深層学習に代表される機械学習の性能が飛躍的に向上したことによるところが大きい．

今から10年程前，2007年1月に，電子情報通信学会料理メディア研究会設立記念シンポジウムが京都大学で開かれた．本稿執筆者の井手や山肩はそれ以前から料理レシピや調理作業映像に関する研究を独立して進めていたが，これを機に国内コミュニティの形成と研究分野としての発展が始まった．今や情報系の国内学会に行けば，食メディアに関するセッションがあることすら珍しくなくなって，皆さんから面白いと言っていただけようになった．また，2009年から毎年開催している国際ワークショップCEA（Workshop on Cooking and Eating Activities）も本年で9回目を迎え，少しずつ海外の研究者も巻き込みながら，国際的なプレゼンスを高めている．

本稿では，このような社会的・技術的背景をうけ発展し

てきた食メディアに関する最先端の研究について，第2節で一般画像認識技術の方法論を料理画像に適用して高精度な認識を実現している柳井が食事画像に関する動向を，第3節で食メディア研究の黎明期から料理レシピ解析に取組んできた山肩が料理レシピに関する動向を，各々紹介する．

2. 料理画像に関する研究動向

深層学習の登場によって画像認識の性能は飛躍的に向上し，1,000種類の物体カテゴリ認識をはじめ多くの認識タスクにおいてすでに人間を上回る性能が実現されていると言われている．食事画像認識は一般物体認識の性能が向上した2010年前後からその応用問題として盛んに研究されてきたが，現在，一般的な料理カテゴリの認識に関しては実用的な精度を達成しており，ほぼ解決済みの問題となりつつある．例えば，Google Photoなどのクラウド写真管理サービスやGoogle Cloud Vision APIなどの画像認識WebAPIサービスでは，一般的なタグに加えて食事カテゴリのタグ付けも画像認識によって自動的に行われるようになっている．また，食事記録サービスアプリのFoodLog appでは深層学習による食事画像認識機能が含まれているなど，食事画像認識は実用的な技術として利用され始めている．一方，研究レベルでは深層学習による料理レシピ検索や画像生成など，新たな取組みも行われている．本節では，研究の歴史と最新の研究動向について簡単に解説する．

2.1 研究の歴史

食事画像認識は，局所特徴とそれに基づくBag-of-Features (BoF) 表現が一般化し，一般物体認識技術が大きく発展した2008年頃から本格的に行われるようになった．当初は，データセットが整備されておらず，研究を行うにはまず自前で食事画像データセットを作成することが必須であったが，2009年のICIPでIntelのChenらによるPittsburgh Fast-food Image Dataset (PFID) が世界初の公開食事画像データセットとして発表された．アメリカの大手ファストフードチェーンのメニューの食品61種類合計1,098枚の食事画像からなる小規模なデータセットであったが，PFIDを用いた研究は多くなされ，ハンバーガーの野菜・パン・肉など食材の位置関係の特徴ベクトルにする手

^{†1} 名古屋大学 大学院情報学研究科

^{†2} 電気通信大学 大学院情報理工学研究科

^{†3} 東京大学 大学院情報理工学系研究科

"Research Trends in Food Media" by Ichiro Ide (Graduate School of Informatics, Nagoya University, Nagoya), Keiji Yanai (Graduate School of Information Science and Technology, The University of Electro-Communications, Tokyo) and Yoko Yamakata (Graduate School of Information Science and Technology, The University of Tokyo, Tokyo)

法など、PFIDに特化した手法なども提案された。同じく2009年には、筆者の研究室からも同じICIPで(しかもポストの場所がPFIDの隣で)食事画像認識の研究発表を行った。その際、主に日本の食事50種類各100枚の合計5,000枚からなる食事データセットとして主にWebから収集した画像セットを構築した。この50種類は現在、食事画像認識研究で広く使われているUEC-FOOD100の一部となっているが、当時はまだ非公開であった。この研究では、BoF、色ヒストグラム、Gabor特徴など9種類の視覚的特徴をMultiple Kernel Learning (MKL)を用いた方法で特徴統合することで、50種類の料理について61.34%の割合で正しく分類することができた。さらに、2010年には85種類に拡張して、特徴量を追加し精度を向上させた。

以上の研究では、認識対象の画像には食事が1種類しか含まれていないことが前提となっていたが、その後、ご飯、味噌汁、おかずが含まれる「定食」のような複数品目の料理画像に対する認識を、Deformable Part Model (DPM)やEfficient Subwindow Search (ESS)などの物体検出手法を組合せることによって実現した。この研究のために、85種類の食事画像データに15種類を追加して100種類、約1万枚とし、さらにすべての画像中の料理1品毎にバウンディングボックスを付与し、UEC-FOOD100として一般に公開した。UEC-FOOD100は、今日、標準的な食事画像データセットの一つとして広く利用されている¹⁾。

食事画像認識は、一般物体認識技術の進歩とともに発展し、2012年前後にはBOFよりも精度が高いとされるFisher Vectorが食事認識に利用された。さらに2014年には深層学習手法であるDeep Convolutional Neural Network (DCNN)が食事画像に適用され、大幅に認識精度が向上した。101種類各1,000枚という大規模食事画像データセットETH-Food101が公開されたこともあり、今日、食事画像認識においても他の一般物体認識と同様に深層学習ネットワークを利用することが一般的となっている。最新の研究では、ResNetの改良版であるWideResNetが適用され、研究開始当初の2009年の手法では認識結果第1位での分類率(Top-1) 51.6%、5位以内(Top-5) 76.8%だったUEC-Food100の100種類食事カテゴリー認識の精度がTop-1 89.6%、Top-5 99.23%まで飛躍的に向上している²⁾。Top-5がすでに100%近くになっており、炒飯とピラフ、オムライスとオムレツなど画像だけでは互いに区別が困難なカテゴリーが含まれていることも考えると、研究的にはこれ以上性能を上げることが難しい段階までできていると言える。

2.2 カロリー量推定

食事画像認識は主に一般物体認識の応用問題として扱われてきたが、海外では医療応用として食事管理のために取り組んでいるグループもある。例えば、米国Purdue大学、スイスBern大学などにおいては、食事管理のために、食事画像からの食事カロリー量や炭水化物などの栄養素量推

定が研究されている。現在、カロリー消費量は腕時計型の運動量計を装着することによってある程度自動に把握することができるが、カロリー摂取量の把握は食事からカロリー量を推定しない限り難しい。もし写真を1枚撮影するだけでそれが可能となれば、カロリー入出力量の管理が容易に可能となり、人間の健康にとってさまざまな有用な応用が期待できることとなる。

食事カロリー量推定には、カテゴリー認識に加えて食事の量の推定が必要となるが、そのための標準的な方法としては、領域分割による2次元による方法と、複数視点画像からの3次元形状復元にに基づく3次元による方法がある。いずれもカテゴリー毎の単位面積もしくは単位体積あたりのカロリー量密度を掛け合わせることによって、カロリー量を推定する。Googleが提案したim2calorieでは、CNNによって単一画像から料理の3次元形状を推定し、カロリー推定を行っている。2015年に発表され、Androidアプリとして提供されることが予告されたが、実際には現在に至るまで提供されていない。実用的な精度を実現できないことが原因と考えられるが、量推定の困難さと、カテゴリー毎に決まった固定カロリー量密度として推定するだけでは不十分であることに起因するものと想像される³⁾。筆者らもCNNを用いたセマンティック領域分割手法によるカロリー量推定を試みている(図1)。ここでは、味噌汁茶碗、ご飯茶碗など大きさの変動が少ない食器に優先順位をつけ、シーン中で最も大きさの変動が小さい食器の大きさとその中身のカロリーを標準サイズ・カロリーとして、それを基準とした面積比とカテゴリーに応じたカロリー比からカロリー量を求めるという簡易的な手法を用いており、100種類の食事に対応している。

一方、東京大学の相澤らから画像検索によってカロリー量推定を行う方法も提案されている⁴⁾。食事画像6,512枚にカロリー量を人手で付与しカロリー量付きの画像データベースを構築し、カロリー量未知の食事画像に関して画像DB検索を行うことによって、カロリー量を推定している。この研究は2011年に行われたため、深層学習は使われておらずBoFなどの古典的画像特徴が利用されていた。そこで、筆者らはCNNによる特徴量を同様の実験を行い、CNNによる回帰によるカロリー量推定と比較したところ、後者の性能の方が高く、特にカロリー量に加えてカテゴリーや食材名など付加情報も同時に学習したマルチタスクCNNが最も性能が高かった⁵⁾。なお、相澤らのカロリー量付き画像データは商用サービスFoodLogの画像を利用しており、非公開であったため、筆者らの研究ではカロリー量付きレシピサイトから8万枚のカロリー量付き画像を収集して利用した。

カロリー量推定には以上のように、量推定を行う方法と、画像から直接推定する方法の2通りがあるが、いずれもまだ研究が充分とは言えず、研究を促進するためにも今後は

図1 CNNを用いた領域分割によるカロリー量推定例

カロリー量付きの大規模食事画像データセットの整備と公開が望まれる。

2.3 最近の動向

現在、研究レベルでは食事カテゴリー認識の精度が充分向上したこともあって、料理の食材認識、画像からの料理レシピ検索、深層学習画像生成技術を用いた食事画像生成など多様な研究が行われるようになりつつあり、新たな研究の広がりを見せている。

最近、料理レシピサイトが一般的になり、大量の画像付き料理レシピ情報を容易に入手できるようになったため、画像による料理レシピ検索の研究が盛んになりつつある。特に、CNNによる画像特徴量と Word2Vec や GloVe などによる分散表現による食材や料理手順の特徴ベクトルを同じ空間に埋め込む (Embedding) することによって、画像とテキストの双方向からのマルチモーダル検索に関する研究が行われている⁶⁾。また、Multi-task CNNによって、食材と料理カテゴリーを同時に推定する研究も行われている⁷⁾。

近年、盛んに研究されている深層学習による画像生成技術である Generative Adversarial Network (GAN) による食事画像の生成の研究も行われている。図2に Conditional GANによるプレーンオムライス画像から文字あり画像への連続画像生成の例を示す。形は変えずに文字属性だけを連続的に変化させることによって、徐々に文字が現れるような画像列を生成することが可能である。将来的には、料理レシピからの完成料理画像や手順動画の自動生成、複数

の料理画像の合成による新しい料理画像の自動生成などの研究に発展させることも期待できる。また、撮影した料理画像をおいしそうに見えるように変換したり、食事量を自由に増減させた画像に変換したりするなど、撮影した料理画像に対して意味的な加工を加えることも今後、可能となるであろう。

3. 料理レシピに関する研究動向

料理は材料を加工し他の材料と組合せて、新しい製品を作る「ものづくり」であり、料理レシピはその手順書である。世界中の多くの人々が、必ずしも専門的な教育を受けることなく、家族や友人から教わったり独学で学んだりしてその技術を習得するばかりか、ときには自ら手順を考案して新しい料理を生み出している。ただ生命を維持するためだけでなく、健康のため、喜びのため、自分だけでなく家族やゲストのためと、懸命に料理を行う人々をICTによって支援することを目的として、料理レシピを活用する情報処理研究が数多く行われている。

情報処理研究の対象となるのは電子的に記述された Web レシピである。現在に続くレシピサイトの多くは、日本を含む先進諸国において一般家庭へのインターネット普及率が10%前後に達した1997年前後に開設されている。1995年に日本初の本格レシピサイトとして大阪ガスの「ボブとアンジー」(現在はオーグス総研が運営)、1997年にキッコーマン「ホームクッキング」、1998年にクックパッドがそれぞれ開設され

図2 Conditional GANによるオムライス画像の連続変化

ている。米国では1995年に料理番組チャンネルFoodNetworkのWebサイト、続いて1997年にAllrecipes.com、1999年にFood.com*1が開設された。

2000年代前半までは専門家が作った料理レシピを公開するスタイルであり、各サイトの料理レシピ掲載数も百から千のオーダーであった。米国のFood Network、NHK「きょうの料理」、「キューピー3分クッキング」、「上沼恵美子のおしゃべりクッキング」など料理番組に付帯したレシピサイトか、あるいはネスレ社がマギーブイヨン、キッコーマンが醤油を使った料理レシピを紹介するというように、自社製品の販促の一環としてレシピサイトを運営する場合が多かった。その後2000年代中頃になると、送り手と受け手が流動化するWeb2.0の波を受けて、一般ユーザが自ら考案した料理レシピを投稿する「ユーザ投稿型料理レシピ」が急増し始めた。このユーザ投稿型料理レシピは今日に至るまで順調に増加を続けており、2017年9月現在、クックパッドは275万件、楽天レシピは137万件に到達している。ちなみに米国では上述のAllrecipes.comとFood.comが2大ユーザ投稿型料理レシピサイトであるが、筆者が2017年4月に調査した時点でアクセス可能なレシピ数はAllrecipes.comが約5.5万件、Food.comが約50万件と、クックパッドや楽天レシピには遙か及ばない。Web解析サービスSimilarWebによれば、2017年7月現在、“Cooking and Recipes in the world”部門においてcookpadが世界総合1位を獲得しており、料理レシピ掲載数においてもユーザアクセス数においても、日本は世界一のWebレシピ大国と言って過言ではない。

料理レシピを対象とする情報処理研究は大きく二つに分けることができる。一つは、料理レシピをシナリオとして人間の調理行動を認識し支援しようとするもの、もう一つは料理レシピ検索・推薦である。前述の通り、歴史的には料理の専門家が作って企業が提供するようなプロの料理レシピが先に登場したことから、それを活用した調理行動支援の研究が2000年前後に始まった。

2000年前後と言えば、ちょうど「ユビキタス」という言葉が広く周知されるようになった頃である。総務省の情報通信白書*2によれば2002年に国内におけるインターネットの人口普及率が5割を超えており、Yahoo! BBがお試しキャンペーンを大々的に展開し、その場でモデムを家に持ち帰ることができたのもその頃。電子レンジや冷蔵庫、ポットなどの家電がインターネットに繋がった「インターネット家電」が登場したのもこの時期である*3。このような時代の流れの中で、家電と連携してセンサや操作の情報

を得たり、システムキッチンにカメラを内蔵させたりして調理状況を把握し、インターネット経由で取り寄せた料理レシピの知識を活用することで、キッチンがまるでカーナビのように調理者を導く調理支援システムが考案された。

調理支援システムにおいて、料理レシピは、調理者に対し調理法を提示するための教本としての役割だけでなく、調理者の調理行動理解のためのシナリオとしても利用された。カーナビのような支援を行うためには、何はともあれ調理者が何をやっているかを知る必要がある。しかし、調理映像を認識するも加工動作は多岐にわたり、同じ「剥く」でも包丁を使う場合とピーラを使う場合とでは動きがまったく異なる。また、食材を認識するも、調理では食材は剥かれて切られて混ぜられて焼かれて、元の食材とはまったく違う物体になることから、同名の物体は何らかの意味で同質であることを前提とする一般的な物体認識手法を適用することが難しい。そこで、調理開始時点から観測データと料理レシピを逐次対応づけることにより、どこまでの手順が完了したかを把握する手法が考案された。観測データとしては、観測映像において食材領域追跡や食材・調理動作認識を行ったり、食材や調理器具にRF-IDタグやARマーカを貼ったりすることで食材や道具の使用や動きを取得する手法等が用いられた⁸⁾。

効率的な調理手順を提案するのも一つの課題である。調理では、パスタを茹でながらソースを作るというように、同時に複数の作業を並列して行わなければならない場面が頻繁にある。一汁三菜のように複数品目調理する場合はなおのこと、すべての料理が同じタイミングで「できたて」の状態のできあがるのは、こと料理の初心者には至難の業である。そこで、料理レシピの手順文書を自然言語処理技術により解析し、作業フローのグラフに変換した上でスケジューリングアルゴリズムを適用することで、調理者に効率の良い手順を提案するシステムが考案された⁹⁾。

2000年代後半になると、前述の通り、クックパッドの料理レシピが爆発的に増加したことから、データ工学分野を中心に、料理レシピ検索の研究が増えてきた。さらに2012年から楽天レシピ、2015年からクックパッドの料理レシピデータセットが、国立情報学研究所を介して研究利用のために提供されたことにより促進され、現在、料理レシピ検索は情報検索の一つのトピックとして多くの研究が発表されている。

料理レシピ検索サイトでは料理名や食材名を検索クエリとして料理レシピを検索するのが一般的だが、ユーザは必ずしも自分が食べたい料理を具体的にイメージできているわけではない。「今日は暑かったからさっぱりしたものが食べたい、でも冷麺はお昼に食べたから違うもので、簡単に作れるものがいい」等、条件は複合的で多様である。

Webレシピのほとんどはユーザが投稿するユーザ投稿型料理レシピであるが、ユーザ投稿型はプロが制作・編集す

*1 Food.comは2017年9月に、Genius Kitchen (<http://www.geniuskitchen.com/>)へと統合された。

*2 平成23年版情報通信白書第4章第1節インターネットの利用動向図表4-1-1-1「インターネットの利用者数および人口普及率の推移」を参照のこと。

*3 1999年9月に市場投入されたシャープのインターネット電子レンジ「インターネットDEこれつくろ!」であった。

図3 レシピの手順文書からのフローグラフ抽出

る料理レシピと違い、いろいろな意味で統一されていないことも問題を難しくしている。例えば、タイトルに「簡単」とある料理レシピは、ない料理レシピに比べて本当に簡単なのか？「本格」とある料理レシピは何をもって「本格」と言っているのか？このような疑問に対し、タイトルに含まれる修飾語と、食材や調理用語との関係を統計的に分析することにより、その名付けの根拠を明らかにする手法が提案されている¹⁰⁾。

一般的な料理レシピ検索では、料理レシピ内で使われている用語の分布のみを考慮するBoW (Bag of Words) 法によって料理レシピを特徴付けるのが一般的だが、より深く活用するためには、「どの食材にどの順番でどの加工を加えたのか」を理解する必要がある。そこで、自然言語処理技術を活用して料理レシピの手順文書からフローグラフを抽出する研究が行われている（図3）¹¹⁾。また、ほとんどのWebレシピには完成した料理の写真が添えられているため、完成写真とレシピテキストの関係を扱う研究も行われている。完成写真と料理名との組合せを教師データとして食事画像認識を行ったり、料理の写真から材料を推定してそれらしい料理レシピを検索したり、辛いのか甘いのかといった大まかな味を推定する試みがある¹²⁾。

料理レシピ研究の国際的な動向としては、事例ベース推論のタスクとして取り上げられる他は、料理レシピ検索・推薦タスクを含めて、日本から発信されるものが多かったが、2015年前後から海外の研究グループによる発表が増えてきているように感じている。この1, 2年では、クラシル (<https://www.kurashiru.com/>) のように、一人称視点で調理の様子を撮影したレシピ動画付きのWebレシピが国内外で急増していることから、これを学習データとして、レシピ動画からレシピテキストを生成するような研究も増えてくると思われる。日本は料理レシピ大国として海外に遅れを取ることなく、質の高い研究成果を発表していくことが期待される。

4. むすび

本稿では、食メディアに関する研究の最先端について、食事画像・料理レシピに分けて紹介した。聖域に入り込むには一段の工夫が必要だが、10年後にはこれらの技術の発

展形が実際に台所で使われるようになっていくことを期待している。

(2017年8月7日受付)

〔文 献〕

- 1) Y. Matsuda, H. Hoashi and K. Yanai: "Recognition of Multiple-Food Images by Detecting Candidate Regions", Proc. 2012 IEEE Int. Conf. On Multimedia and Expo, pp.25-30 (July 2012)
- 2) N. Martinel, G.L. Foresti and C. Micheloni: "Wide-Slice Residual Networks for Food Recognition", arXiv Preprints, arXiv:1612.06543 (Dec. 2016)
- 3) A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang and P.K. Murphy: "Im2Calories: Towards an Automated Mobile Vision Food Diary", Proc. 15th IEEE Int. Conf. On Computer Vision, pp.1233-1241 (Dec. 2015)
- 4) T. Miyazaki, G.C. de Silva and K. Aizawa: "Image-based Calorie Content Estimation for Dietary Assessment", Proc. 2011 IEEE Int. Symposium on Multimedia, pp.363-368 (Dec. 2011)
- 5) T. Ege and K. Yanai: "Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions", Proc. 25th ACM Int. Conf. On Multimedia, Thematic Workshop on Understanding (Oct. 2017)
- 6) A. Salvador, N. Hynes, Y. Aytaç, J. Marin, F. Ofli, I. Weber and A. Torralba: "Learning Cross-modal Embeddings for Cooking Recipes and Food Images", Proc. 2017 IEEE Conf. On Computer Vision and Pattern Recognition (July 2017)
- 7) J. Chen and C.W. Ngo: "Deep-based Ingredient Recognition for Cooking Recipe Retrieval", Proc. 24th ACM Int. Conf. On Multimedia, pp.32-41 (Oct. 2016)
- 8) 山肩, 船富, 上田, 辻, 美濃, 中内, 宮脇, 中村, 椎尾: "料理を作る (小特集「生活に役立つメディア処理 - 料理行動を科学する -)", 信学誌, 93, 1, pp.39-47 (Jan. 2010)
- 9) R. Hamada, J. Okabe, I. Ide, S. Satoh, S. Sakai and H. Tanaka: "Cooking Navi: Assistant for Daily Cooking in Kitchen", Proc. 13th ACM Int. Conf. On Multimedia, pp.371-374 (Nov. 2005)
- 10) 高橋, 小山, 大島, 田中: "投稿型レシピサイトにおけるレシピ名の修飾表現の適合性判定手法", 信学論, J94-A, 7, pp.467-475 (July 2011)
- 11) 前田, 山肩, 森: "手順文書からの意味構造抽出", 人工知能学論, 32, 1, pp.1-8 (Jan. 2017)
- 12) 松長, 道満, 平山, 井手, 出口, 村瀬: "料理画像および素材一覧に基づく料理の食味表現推定", 信学技報, MVE2014-82 (Mar. 2015)

井手 一郎 2000年、東京大学大学院工学系研究科博士課程修了。同年、国立情報学研究所助手。2004年、名古屋大学大学院情報科学研究科助教授。2007年、同准教授となり、現在に至る。博士(工学)。正会員。

柳井 啓司 1997年、東京大学大学院工学系研究科修士課程修了。同年、電気通信大学情報工学科助手。2006年、同大学助教授、2007年、同大学准教授、2015年、同大学大学院情報理工学研究科情報学専攻教授となり、現在に至る。博士(工学)。

山肩 洋子 2005年、京都大学博士課程満期修了。同年、同大学研究員。2006年、(独)情報通信研究機構専攻研究員。2010年、京都大学大学院情報学研究科特定講師。2011年、同大学准教授。2015年、日本学術振興会特別研究員(RPD)となり、現在、東京大学大学院情報理工学系研究科において、研究に従事。博士(情報学)。