

食事画像認識による食の情報共有・可視化支援システムの提案 Proposal of Food Log Sharing and Visualization Support System based on Food Image Recognition

大井 翔[†] 堂土 奨[†] 佐野 睦夫[‡]

Sho Oi Sho Dodo Mutsuo Sano

1. はじめに

近年, Food Log[1]や食事管理のインターネットサービスなど, 健康を意識した食事に着目したアプリケーションが実装されている. これらのサービスでは, ユーザが何を食べたかのログを残し, 食生活の改善を行うことができる. 食事のログを残すことで栄養のバランスを知ることが可能であり, 接種不足や過剰摂取などを確認することができ, 食生活の改善には食事の内容を記録することが効果的である[2-5]. これらの研究では, 食生活を改善するために食事ログを用いて調理レシピを提示するケースである.

しかし, 食事ログを残すことは手間がかかる場合が多い. そこで, 画像処理を用いることで, 簡単に記録する研究や[6, 7], 撮影した食事画像から食事バランスの推定を行う研究もある[8]. このように食事画像から自動的に認識するシステムの提案がされている.

一方, 食事に対してはCookPad[9]やTwitterなどのユーザの投稿型のサイトでは, 食事に対する情報を自由に投稿している. 特にCookPadであれば自分のレシピをおいしそうに表現している[10]. 食事において「おいしさ」や「甘い」などの言葉の情報は重要であり, これらの示す言葉から人々がイメージする食品が何であるか理解する.

本研究では, 調理において言語情報も重要であると考え, ユーザが調べるだけではわからないような言語の情報を提供することで食に対してより一層興味を持ってもらうことを目的とする. 研究の課題として, ①食事画像に対する認識, ②食事画像に対する言語処理である.

食事画像に対する認識は, 食事画像をスマートフォンなどのデバイスで撮影し, 撮影した食事画像に対して一般物体認識・深層学習による認識それぞれを行い, 認識の結果から何の料理であったか推定する.

言語処理としては, 料理がメディアによってどのような差が見られるのかを分析すると同時に料理に含まれている情報や感情など相関性を示し, ユーザに調べるだけでは分からない言語情報を提供する. インターネット上の2つのメディアを対象とし, 比較分析を行い, 頻度を元にユーザに情報提供を行う. 具体的には, Twitter, ユーザ投稿型レシピCookPad, 料理画像に対する個人の潜在能力を問うアンケートを行い, それらの違いを分析する.

2. 関連研究

ここでは, 関連する研究について紹介する.

2.1 食事関連アプリケーション

現在, 実装されている料理に関するアプリケーションとしてFoodLogがある. FoodLogは, 食事の写真を撮って設

定したメールアドレスに送信することで, 食事に関する様々な情報を管理することができるWebサービスのことである. 食事日記としての使い方や, 健康管理のツールとして使い方, また他の人の公開されている食事を見ることもできる. また, アップロードされた写真から「主食」「副菜」「主菜」「牛乳・乳製品」「果物」に分類し食事バランスを自動測定する機能がついており, 一定期間のバランス推移やグラフで見ることができる. つまり, 次の食事や翌日のメニュー選択にも役立つことができる. 位置情報記録できるため, 食事マップ作成可能である. 公開されている他の人の食事写真は「年齢」「性別」「タグ」などで検索ができるため, 他の人のメニューを参考にできる[7].

一方, CookPadでは, ユーザが自分の料理レシピを写真と共に投稿したり, 他のユーザのオリジナルレシピを参照して献立の参考にするを目的としてWebツールである. FoodLogと違い, 食事管理という機能というよりレシピ共有を目的としている. また, 関連研究として柳井らによるスマートフォン上でのリアルタイム食事画像認識アプリ実装が行われている. そこでは, 食事にスマートフォンのカメラをかざすだけで, 連続的に食事認識を行い, 食事の記録が容易に行えるシステムである. また, 食材を連続的に認識し, その食材に関係した料理のレシピを次々に推薦してくれるシステムである[8]. 以下に図2.1料理関連アプリケーションを示す.

2.2 食事画像認識

一般物体認識としてVisual Wordsを用いたBag of Features (BoF)と呼ばれる方法がある[11][12]. これらは局所特徴量からVisual Wordsを作成し, 認識をする方法である.

調理に限定した認識として, 松田らは, 機械学習の方法であるMultiple Kernel Learningを用いて50種類の食事画像データの分類を試みている[13][7]. また, 1枚の画像中において複数種類の料理画像が含まれている場合の認識も行っている[14]. また, 数藤らは料理に使用されている食材を学習することで精度の向上を目指している[15]. 我々もこれまでに, BoF特徴量ベースでクラスタリングをする際にk-means法を用いた方法[16]や混合ディレクテ過程に基づき自動クラスタリングを行う方法[17]をしてきた. しかし, 通常これらの従来方式では認識精度は7~8割程度と言われており, パラメータや特徴量を自分で決定する必要性があった.

特に, 混合ディレクテ過程に基づく方式はあまり用いられることが少なかった. 理由としては深層学習が登場したためである. 深層学習は, 学習時間はかかるものの, 認識にかかる実行時間が短く行え, 精度も9割に見込む認識率を誇るといわれている. しかしながら, パラメータやデータセット, ネットワークの構築に苦戦すると言われている.

[†]大阪工業大学 大学院情報科学研究科

[‡]大阪工業大学 情報科学部

我々の研究[16][17]では、初めに Bag-of-Features を導入し研究を行ったが 4 割程度の精度しか得られなかった。また、食材推定にも試みたが、料理画像では、食材が変化しやすく、形状や色を特定するのが困難であり、認識精度は 5% 向上したが、料理画像が増えるに従い特定が困難であった。そして、現在、深層学習を取り入れ、パラメータやデータセットを変更し認識率向上に励んでいる。

2.3 食の言語処理

近年、形態素解析を行えるツールとして MeCab が流行し、Python や Ruby などから実行し単語の頻度を計測することが可能となった。また、感情分析なども行うことができる。言語情報を抽出するフィールドとして、Twitter や Web 抽出や CookPad など利用されている。関連研究として、加藤らの研究では、おいしさを表す言葉着目し、おいしさの表現は感想、宣伝、紹介などの目的によって使い分けられ、Web サイト、投稿型レシピサイトなどのメディアによっても異なると仮説を立てている。そこから、インターネットのメディア上でのおいしさを表す言葉であるシズルワードと食品との関係の比較分析を行う研究を行った[10]。我々は、これを参考とし、ある料理名に対し Mecab による頻度分布を行い、ネット上の情報と個人の情報を差分しユーザーに新たな情報を提示し想起・発展を促すシステム作成に通じ組んでいる。

3. 食の情報共有・可視化システム

本システムは、①画像認識部（ボトムアップ）、②言語処理部（トップダウン）の 2 種類の処理部から構成されている。画像を認識し、言語情報を出力すると同時に、言語情報から画像の推定を行う。トップダウン・ボトムアップ制御に基づく図 1 のような構成で処理を行う。

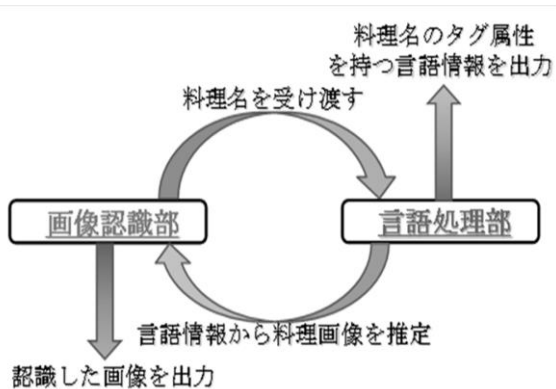


図 1 システム構成図

画像認識部分では、スマートフォンを用い、撮影された画像をサーバへソケット送信し受け取る。その画像を予め学習が施されているモデルと比較し、最も尤度の高い料理名をユーザーへ言語データとして出力する。言語出力部分では、予め形態素解析を行い頻度算出し定義ワードのみ用いたネット上の分析結果を用いる。これとユーザーのために作った個人アンケートにより引き出された料理画像に対する潜在的な意見を比較し、ユーザーへ想起・発見を促す。処理フローの詳細を図 2 示す。

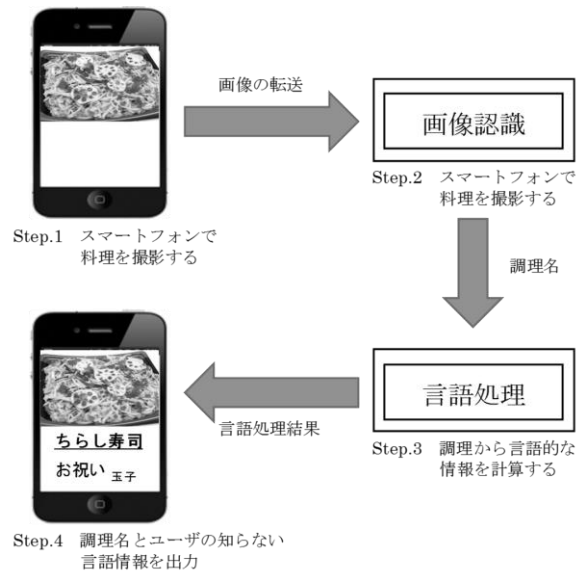


図 2 システムの処理フロー

本研究では、通信の部分は現在未実装で開発中であり、画像認識部、言語処理部それぞれに対して行った処理結果を示す。

3.1 画像認識部

ここでは、これまでに我々が行ってきた一般物体認識の手法と深層学習の手法を述べる。

3.1.1 一般物体認識による料理画像認識

一般物体認識による調理画像認識では、初めに画像からカラーヒストグラムを算出する。カラーヒストグラムは HSV 色空間での Hue を用いた。カラーヒストグラムを用いて類似している料理名の上位 5 種類を絞る。これは、料理画像にはいろいろな見え方などがあるが、大きく色の特徴量が異ならないと考えて行っている。

次に、SIFT 特徴量ベースの BoF を用いる。ここで、Visual Words を作成するためにクラスタリングをする必要があるが、今回混合ディリクレ過程を用いてクラスタリングした。図 3 に Visual Words の作成手順を示す。

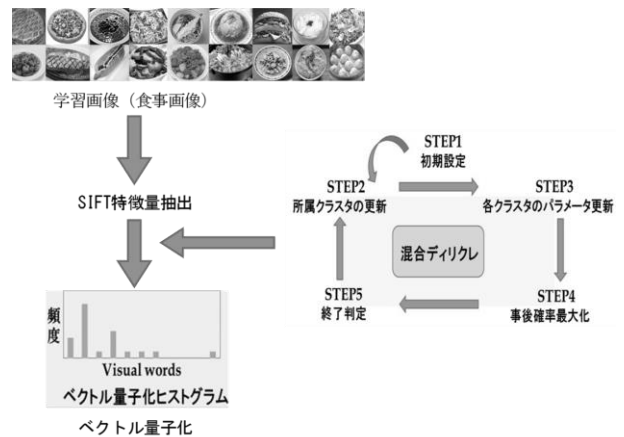


図 3 混合ディリクレ過程を用いた構成図

全体の処理の流れは次のとおりである。

<学習の工程>

- STEP1 各料理リストを作成
 STEP2 料理画像から SIFT 情報量を抽出
 STEP3 料理画像の SIFT 情報量を K-means 法で学習(ベクトル量子化)
 STEP4 混合ディリクレにより 2500~10000 個のクラスタ数の中から最適なクラスタ数を算出
 STEP5 料理画像の平均カラーヒストグラム算出

<認識の工程>

- STEP1 未知画像を入力
 STEP2 未知画像から色情報を抽出
 STEP3 平均カラーヒストグラムのデータと比較し、料理名の候補を 20 個から 5 個まで絞る
 STEP4 未知画像から SIFT 情報を抽出(ベクトル量子化)
 STEP5 料理画像全体の SIFT 学習データと比較し料理名の候補を出力し最も認識率が高いものを入力画像の料理名とする

3.1.2 深層学習による料理画像認識

深層学習は、画像などのデータを入力すると、情報が第 1 層からより深く伝達されるうちに、各層で学習が繰り返される。この過程で、これまでは画像や音声などそれぞれのデータの研究者、技術者が手動で設定していた特徴量が自動で計算される。深層学習を扱うために Caffe や Tensorflow などの様々なライブラリがあるが、本研究では Chainer と呼ばれるライブラリを用いた。本稿で使用したネットワークとして、図 4 に示すような AlexNet を用いる。

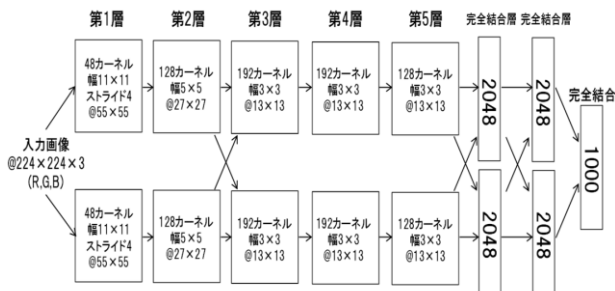


図 4 AlexNet の構成

3.2 言語処理部

料理における一般の言語情報として、スマートフォンを用いることで、調理映像が取得された日付情報、位置情報が簡単に取得できる。また、Twitter は手軽に投稿することができ、ユーザの率直な意見や気持ちを投稿している。また CookPad も同様にユーザ視点のレシピ投稿のため、ユーザの意見が反映された文章になっていると考えられる。そこで、食事情報にユーザの感情などを表現しているデータを用いる。しかし、Twitter や CookPad から抽出した場合、余計な情報が多く含まれるため本研究では、以下のような 6 種類の定義属性を設けた。

- ・料理名
- ・料理に使われている食材
- ・料理発祥の国名
- ・風味 (甘い・辛い・すっぱい・苦い・塩辛い)
- ・値段 (高い・安い)
- ・感触 (固い・柔らかい)

6 種類の情報を取得する手順を以下に示す。

- (1) 各料理名を含む Twitter 上・CookPad からテキストデータ一文を抽出する
- (2) MeCab で形態素解析を行う
- (3) 目視で料理名、国名、風味、値段、感触のそれぞれに属する頻度が最も高いものを選ぶ
- (4) 目視で食材を探し、頻度の高い上位 4 つまでを選ぶ

次に料理における特定の人物の言語情報について述べる。例えば、日記など特定の人物が記しているデータが適切であると考えられるが、抽出するのが良いが難しいため、今回 Google フォームでアンケートを取り活用することにした。アンケートの内容は、先ほどの属性を聞く内容となっており、表 1 に以下に示す。

表 1 アンケートに用いた料理画像

Q1. 料理画像 1 の料理名は何でしょうか？
Q2. 料理画像 1 に使われている食材は何でしょうか？使われていると思う食材を 4 つまで記述して下さい。
Q3. 料理画像 1 は、どの国の料理だと思いましたか？
Q4. 料理画像 1 を見て、どの風味が近いと思いましたか？
Q5. 料理画像 1 を見て、値段は安そうですか？高そうですか？
Q6. 料理画像 1 を見て感触は、柔らかそうですか？固そうですか？

このとき、図 5 に示すような料理の画像のみを見せて答えさせた。



図 5 アンケートに用いた料理画像

一般的な言語情報と特定の人物の言語情報を比較し、差分をとり、該当しなかったものだけをユーザに出力する。例としてそれぞれ選ばれた言語情報を図 6 に示す。

選ばれた言語情報の出現頻度を見やすくするために、本研究では「Word-Cloud」を用いた。文章中で出現頻度が高い単語を複数選び出し、その頻度に応じた大きさと図示する手法のことである。ウェブページやブログなどに頻出する単語を自動的に並べることを指すことを目的としており Python などで実行可能である。文字の大きさだけでなく、色、字体、向きに変化をつけることで、文章の内容をひと目で印象づけることができる。MeCab をインポートすることで、入力したテキストファイルに対し形態素解析を行い、頻度を算出し高いものから順にビジュアル化することができる。また、パラメータを変更することで、単語のランクを利用可能となったり、マスクを設定できたり、表示する単語の最大数を指定、ストップワードを集めて指定色・大きさの変更が可能となる。Word-Cloud を用いた結果を出力結果の一例を次のページに図 7 に示す。

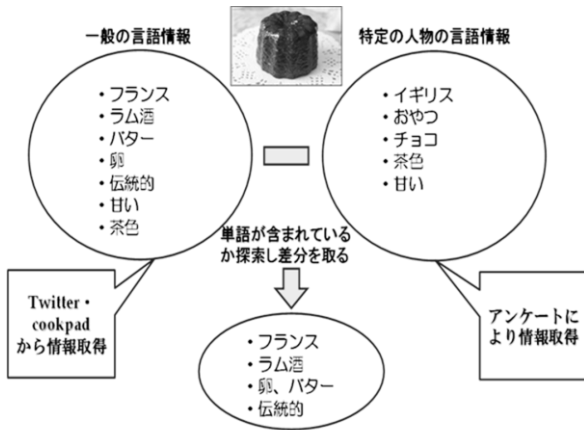


図6 言語情報の差分処理



図7 言語情報の差分処理

4. 実験

初めに、本研究で使用した機器として、グラフィックが「NVIDIA® GeForce® GTX960M」であり、CPUが「インテル®i7-4910MQ」、メモリが16GBのWindows10のパソコンを用いた。

4.1 画像処理部

調理画像のデータベースとして、様々なデータセットがあるが、今回我々は、MyDataSet (OIT-Food-20)を含めて以下に示す4種類のデータセットを用いた。

- (1) OIT-Food-20 (1カテゴリ: 50~100枚, カテゴリ数: 20)
- (2) ETH-Food-101(1カテゴリ: 1000枚, カテゴリ数: 101)
- (3) UEC-Food-256(1カテゴリ: 100~200枚, カテゴリ数: 256)
- (4) VIREO-Food-172(1カテゴリ: 500~1000枚, カテゴリ数: 172)

それぞれのデータセットの詳細として、(2)の「ETH-Food-101」はスイスのチューリッヒ工科大学 (ETHZ) によって作られたものであり、Foodspotting アプリの画像を取り込み構成されたものである。(3)の「UEC-Food-256」は、電気通信大学の柳井らにより作られたものであり、このデータセットのほとんどの食品カテゴリは、日本をはじめとする国で人気のある食品である。このデータセットは、日本で使用されることを意図した実用的な食品認識システムを実装するために構築された。(4)の「VIREO-Food-172」は、香港大学で作られたもので食材の定義まで施されてい

る。同意書を申請することで入手することが可能である。本研究で最も1カテゴリの枚数が多いのも特徴である。

深層学習「Chainer」方式では、料理画像の各データセットの内7割を学習用の画像として使用し、残りの3割の画像をテスト用の画像(未知画像)として使用する。

混合ディリクレ過程を用いた認識については、自動でクラスタ数が推定されるので出力されたクラスタ数に対して認識を行う。さらに、20種類×10枚の合計200枚中何枚正解したかで認識率を計算する。また、この2つの方式の比較は、OIT-Food-20のみで行うものとする。残り3つのデータセットの比較は深層学習「Chainer」方式のみで行うものとする。以下に詳しく比較対象を図8で示す。

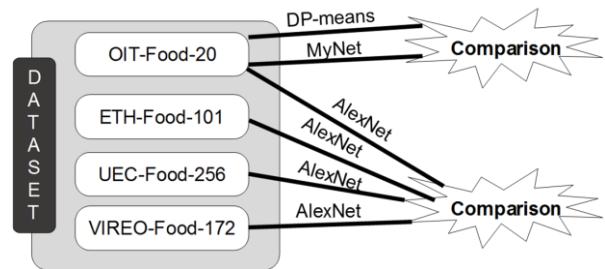


図8 認識方法の比較対象

本稿で扱ったパラメータを以下に示す。

(i) 混合ディリクレ過程 (DP-means)

- ・画素数 256 px × 256 px
- ・事前分布のパラメータ $\gamma A = 0.2$
- ・事後確率判定 $\gamma B = 0.7$
- ・反復回数 250~1000
- ・クラスタ数は 2500~10000

(ii) 深層学習 (AlexNET)

- ・画素数 256 px × 256 px
- ・第8層まで行う
- ・最大値プーリングを使用
- ・畳み込み層とプーリング層は3層ずつ交互に重ねた
- ・並列実行回数2 学習回数 250~1000

4.2 言語処理部

はじめに、一般的な情報として、TwitterとCookPadから情報を得た。それぞれの内訳は以下の通りである。

- ・Twitter: 1種類の料理に対して1,500件
- ・CookPad: 1種類の料理に対して10件

Twitterは「1500 ったー」ツールを用いて抽出を行い、CookPadは手動にて抽出した。

次に、特定の人物に対する情報として、アンケートを対象人数25名(男子17名, 女子8名)に対して行い、情報を得た。今回、可視化ツール (Word_Cloud) を用いているが、OIT-Food-20にある20種類のデータに対しても時間がかかるため、その中でもランダムに10種類の料理「カオマンガイ、カヌレ、天津飯、パニーニ、ミネストローネ、棒棒鶏、パッタイ、ラタトゥイユ、酸辣湯、ブリュレ」を対象にして、言語処理を行った。

5. 実験結果

画像処理に関する実験と言語処理に関する実験について述べる。

5.1 画像処理部

混合ディレクレ過程を用いた OIT-Food-20 に対する Test 画像の認識率, 計測時間を表 2 に示す. 深層学習(MyNet)を用いた OIT-Food-20 に対する Train 画像・Test 画像の認識率, 計測時間を表 3 に示す. 深層学習(AlexNet)を用いた ETH-Food-101, UEC-Food-256, VIREO-Food-172 に対する Train 画像・Test 画像の認識率, 計測時間を表 4, 表 5, 表 6 に示す.

表 2 DP-means を用いた OIT-Food-20 の結果

学習回数(回数)	250	500	750	1000
Test 画像認識率(%)	26	35	45	48
計測時間(h)	8.1	16.2	24.3	32.4

表 3 深層学習を用いた OIT-Food-20 の結果

学習回数(回数)	250	500	750	1000
Train 画像認識率(%)	49	65	71	74
Test 画像認識率(%)	51	57	58	59
計測時間(h)	0.5	1.0	1.5	2.0

表 4 深層学習を用いた ETH-Food-101 の結果

学習回数(回数)	250	500	750	1000
Train 画像認識率(%)	22	38	44	50
Test 画像認識率(%)	21	34	38	45
計測時間(h)	21.4	42.8	64.3	85.7

表 5 深層学習を用いた UEC-Food-256 の結果

学習回数(回数)	250	500	750	1000
Train 画像認識率(%)	2.3	7.6	8.8	9.1
Test 画像認識率(%)	1.1	4.2	6.6	7.3
計測時間(h)	8.6	17.2	25.8	34.3

表 6 深層学習を用いた VIREO-Food-172 の結果

学習回数(回数)	250	500	750	1000
Train 画像認識率(%)	37	39	42	47
Test 画像認識率(%)	36	38	41	45
計測時間(h)	21.9	43.8	65.6	87.5

次に, 図 8 で示した比較をグラフにまとめる. 図 9 に混合ディレクレ過程と深層学習の比較を, 図 10 に深層学習によるデータセットの比較を示す.

OIT-Food-20 のデータセットに関して, 未知データに対する認識率 (Test 画像) は, 混合ディレクレ過程の結果より深層学習を用いた結果の方が平均で 18%程度向上していることが分かる. また処理速度は 1/20 という速さになっている.

次に, 学習データごとの認識では, OIT-Food-20 の精度が最もよく 64.75%であった. 処理時間に関しては, 画像のデータ数が異なるため, 平均時間を画像の総枚数で割った 1 枚当たりの時間として算出した場合, ETH が最も早く 0.25(s)であった.

混合ディレクレ過程と深層学習のOIT-Food-20での比較

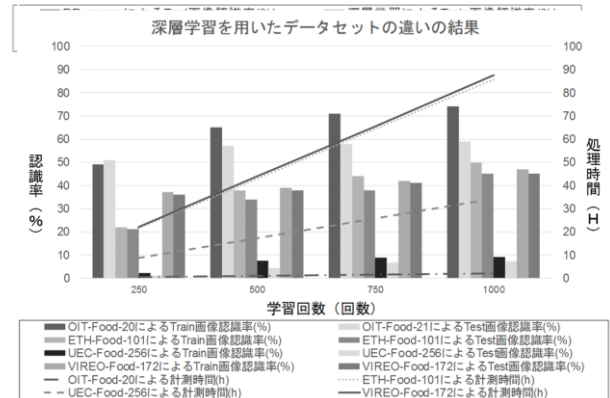
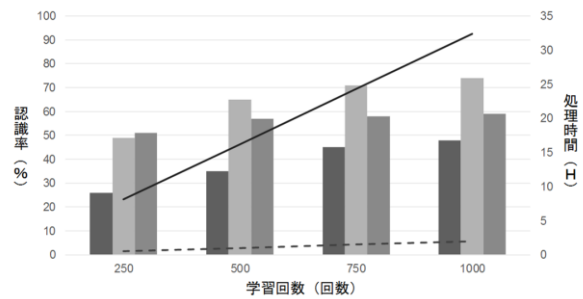


図 9 認識方法の比較対象

5.2 言語処理部

10 種類の料理に対して選出されたものを以下に表 8 として示す. それぞれ「料理名・国名・食材 4 種・風味・値段・感触」の構成となっており, 頻度の高いものから順に選出手動と目視で行った.

次に, 実際にユーザーに気付かせ, 促しを与え情報を提示するときの出力した言語情報である. 図 10 にある男性の結果の一部を示す. 左上の画像が「一般の言語情報」, 向かって右上が「特定の人物の持つ情報」, 向かって下の画像は「この 2 つを差分した情報になる」つまり, ユーザーに教えるべき情報を示す.

6. 考察

画像処理に関する実験と言語処理の考察について述べる.

6.1 画像処理部

はじめに, 混合ディレクレ過程と深層学習による比較だが, 深層学習の方が 11%も認識率が良く, 計測時間もおよそ 8 倍近く深層学習の方が使えることから, 深層学習を用いた方が良いということが分かった. 学習回数が増加するほど精度が上がっているが, 3000 回行った場合も 59%から精度は上がらなかった. つまり, ある程度で収束しているためであり, これ以上の精度向上は期待できないと考えられる.

次に, 学習データベースの比較だが, 認識率がわずか 9%しか得られなかった UEC-Food-256 について考える. このデータベースはほかのデータベースと異なり, 食事以外の余計な部分も画像中に含まれている. そのため, 認識精度が 1 ケタになったと考えられる. また, ETH-Food-101

表 7 10 種類の料理の選出された単語リスト

調理名	カオマンガイ	カヌレ	天津飯	パニーニ	ミネストローネ	パッタイ	ブリュレ	ラタトゥイユ	酸辣湯	棒棒鶏
国名	タイ	フランス	日本	イタリア	イタリア	タイ	フランス	イタリア	台湾	中国
食材 1	パクチー	牛乳	ご飯	パン	タマネギ	エビ	佐藤	パブリカ	うどん	きゅうり
食材 2	鶏肉	バター	カニカマ	チーズ	トマト	もやし	牛乳	チーズ	酢(黒酢)	鶏肉
食材 3	ごはん	蜜蝋	ネギ	トマト	ベーコン	パクチー	生クリーム	トマト	ラー油	もやし
食材 4	商が	ラム酒	玉子	レタス	じゃがいも	焼きそば	卵黄	ズッキーニ	そば	ゴマダレ
風味	辛い	甘い	甘い	辛い	辛い	辛い	甘い	酸っぱい	酸っぱい	辛い
値段	高い	高い	高い	高い	安い	安い	高い	高い	高い	安い
感触	柔らかい	柔らかい	柔らかい	固い	柔らかい	柔らかい	柔らかい	柔らかい	固い	柔らかい

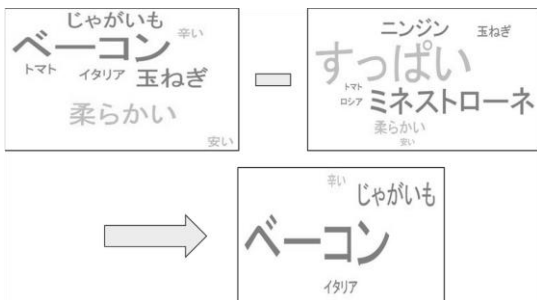


図 10 ミネストローネの言語出力結果

及び VIREO-Food-172 も OIT-Food-20 ほど精度は高くはなかったが、同じような傾向が見られた。この 2 種類のデータセットは、OIT-Food-20 と比べて画像数が多いため、時間がかかる。認識率が 250~1000 回を終えたところで、大きく向上していることから、今後、学習回数を増やすことで他のデータセットよりも認識率向上の見込みがあると考えた。これについては、次回以降の実験で確かめる必要がある。個人的な考察として、データセットにより学習回数の臨界点が違い、その臨界点を導き出すことが今後の課題でもありと考へた。

6.2 言語処理部

一般的な言語情報として、Twitter や CookPad では感情的な用語は手に入りやすいが間違っただけの情報も手に入りやすい。例えば、天津飯は中国の食べ物であるが、日本という単語が多かったために間違っただけの情報を提供してしまっている。また、Twitter では、場所や知覚や味覚など直観的なワードが多いのに対し、CookPad では食材の情報が多く手に入ることが分かった。目視と手動で行っているため、そこを自動でできるように今後取り組みたいと考えている。また、ユーザが出力結果を見てどう感じたのかアンケートによりフィードバックを取る必要がある。

7. おわりに

現状では、画像の認識結果は 6 割程度であるため、実験回数を増やす必要がある。2 種類の方式で比較し、認識精度の結果から深層学習を用いることが望ましいことが分かった。今後は、より精度の高い結果が得られるようにネットワークの構成について調整していく予定であり、他のデータベースにおいても十分な精度を持つモデルを作ることが重要であると考えた。また、主目的であるスマートフォンとの結合部分が完成していないため、画像の通信などの実装も今後行っていく予定である。

謝辞

共同研究をして頂いた京都府立大学の方々、アンケートにご協力いただいた学生の皆さんに感謝の意を表します。

参考文献

- [1] FoodLog, <http://www.foodlog.jp/> (閲覧日: 2017.06.25).
- [2] 苺米志帆乃, 藤井敦: “食事ログにおける栄養バランスを考慮した食生活支援システム”, 情報処理学会 全国大会講演論文集 第 72 回(「情報爆発」時代に向けた新 IT 基盤技術), pp. 191-192, 2010.
- [3] 苺米志帆乃, 藤井敦: “栄養素等摂取バランスの分析に基づく食生活支援システム”. 日本データベース学会論文誌, Vol.8, No.4, pp.1-6, 2010.
- [4] 苺米志帆乃, 藤井敦: “食生活支援システムにおける食事ログとアクセスログの分析”, 第 2 回データ工学と情報マネジメントに関するフォーラム, 2010.
- [5] 苺米志帆乃, 藤井敦: “料理レシピの推薦と栄養バランスの可視化による食生活支援システム”, WebDB Forum 2009, 2009.
- [6] 甫足創, 松田裕司, 柳井啓司: “候補領域推定による複数品目に対応した食事画像認識”, 画像の認識・理解シンポジウム (MIRU2011) 論文集 2011, pp. 234-240, 2011.
- [7] 松田裕司, 甫足創, 柳井啓司: “候補領域推定に基づく複数品目食事画像認識”, 電子情報通信学会論文誌 D, Vol. J95-D, No. 8, pp. 1554-1564, 2012.
- [8] 相澤清晴, 小川誠: “食とコンピューティング: 3. マルチメディア食事記録と画像処理による食事内容解析”, 情報処理, vol.52, No.11, pp.1382-1387, 2011.
- [9] cookpad, <https://cookpad.com/> (閲覧日: 2017.06.25)
- [10] 加藤大介, 宮部真衣, 荒牧英治, 瀧本明代: “インターネット上のメディア毎の「おいしさ」表現比較分析”, DEIM Forum2015, 2015.
- [11] 田中雄翔, 滝口哲也, 有木康雄: “ウェブ画像を用いたカテゴリ別 Visual Words による未知物体判別”, 画像の認識・理解シンポジウム (MIRU2012), 2012.
- [12] 柳井啓司: “一般物体認識の現状と今後”, 情報処理学会論文誌: コンピュータビジョンとイメージメディア, 48, pp. 1-24 2007.
- [13] 上東太一, 甫足創, 柳井啓司: “Multiple Kernel Learning による 50 種類の食事画像の認識 (パターン認識応用, 特集 画像の認識・理解論文)”, 電子情報通信学会論文誌. D, 情報・システム, 93, 8, pp. 1397-1406, 2010.
- [14] 松田裕司, 甫足創, 柳井啓司: “複数品目が含まれる食事画像の認識における共起関係の利用”, 電子情報通信学会論文誌 D, Vol. 96-D, No. 8, pp. 1724-1730, 2013.
- [15] 数藤恭子, 村崎和彦, 神谷叔季, 谷口行信, “料理画像の素材セグメンテーションに基づく成分推定”, 電子情報通信学会誌, No.468(IMQ), No.469(IE), No.470(MVE), pp.73-76, 2014.
- [16] 堂土奨, 佐野睦夫: “食材学習を用いた料理画像認識”, 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎研究会, vol.114, no.485, pp.159-160, 2015.
- [17] 堂土奨, 佐野睦夫: “料理画像認識の高精度化”, 2015 年度 情報処理学会関西支部 支部大会 講演論文集, 2015.