

位置情報付き Twitter 画像を用いた世界の食事傾向分析

岡本 開夢[†] 柳井 啓司[†]

[†] 電気通信大学情報理工学部総合情報学科 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: [†]okamoto-ka@mm.inf.uec.ac.jp, ^{††}yanai@cs.uec.ac.jp

あらまし Twitter は現在、世界中で普及しておりそれに伴って多くの画像が Twitter 上に投稿されている。Twitter 上の画像のなかでも食事画像は、Twitter 上に投稿されている画像種類の大きな要素の一つである。また、この食事画像は地域傾向があるとされており、Twitter のログデータをもとに画像の位置情報からこの食事画像を分析することで、地域傾向における新たな発見をする可能性がある。そこで、本研究では、まず、Twitter 上の画像とログデータから位置情報付き食事画像を抽出し、位置情報から 9 つの地域に食事画像を割り当てる。その後、食事画像の CNN 特徴量からクラスタリングを行うことで 17 料理カテゴリに分類し、料理カテゴリと 9 つのうち 6 つの地域をもとに分析することで、地域別の食事傾向と食事画像別の地域傾向の発見を行った。また、食事画像特徴量を可視化させることで地域特徴が表れている箇所の発見を行った。

キーワード Twitter, 食事画像, CNN 特徴量, クラスタリング

1. はじめに

SNS は現在、世界中で普及しており、それに伴って多くの画像が SNS 上に投稿されている。SNS のなかでも、特に Twitter は若者を中心に普及しており、ツイートとともに多くの画像が投稿されている。また、Twitter にはログデータが存在しており、投稿されたツイートの文章、画像、時間、場所といったメタデータを取得できることから、長野ら [1] の Twitter 画像の地域的傾向分析といった多くの分析に使用されている。さらに、Twitter 画像のカテゴリのなかで食事画像は地域的傾向があるとされている。日本では白米の画像が多く、アメリカでは油が多く使われた肉料理の画像が多いと予想することができるが、それらがどのような地域的傾向があるのかは明確になっていない。そこで、本研究ではこのような地域的傾向を発見するため、Twitter のログデータから取得した位置情報をもとに、食事画像の分析を行う。位置情報付き食事画像の分析を行う際に、優れた表現力をもつ CNN 特徴量を用いる。CNN 特徴量は、ネットワークの中間層から抽出した特徴量のことで、物体の形状や模様を幅広く表現できるとされている。CNN 特徴量を用いた研究は多くなされており、画像認識、分析や生成といった幅広く活用されている。また、分析を行う際に本研究では、テキスト情報を使わずに、画像特徴のみで分析を行っていく。これは、画像とともにツイートに付随するタグといったテキスト情報が、必ずしも適切に画像を表現しているとは限らない。場合によって、画像とはまったく関係のないテキスト情報が付随しているからである。

分析を行う項目は、次の項目の通りに分析を行っていく。

- (1) 地域ごとの食事傾向分析
- (2) 代表的な料理の地域分布の分析
- (3) 食事の地域特有の特徴を表す箇所の可視化

地域ごとの食事傾向分析では、地域ごとにおける上位 5 項目の料理の割合を求め、この地域における料理の割合を地域ごと

に比較することで、地域ごとの食事傾向の発見を行う。代表的な料理の地域分布の分析では、料理ごとに地域割合を求め、この料理ごとの地域割合を観察することで、どの地域がどれだけその料理に対して影響を与えるのの発見を行う。最後に可視化では、画像特徴量をもとに料理画像において画像のどの部分が地域特有な箇所の発見を行う。

2. 関連研究

2.1 地域分析

現在、様々な SNS やインターネット上でユーザーが画像データを投稿しており、これらの画像をもちいた研究が行われている。先行研究として、これらの画像を用いて地域分析を行った、研究として長野ら [1] と Chang ら [2] の研究がある。

長野ら [1] の研究では、テキスト情報を用いずに Twitter 画像の特徴量のみを用いて、クラスタリング分析を行うことで、各地域における画像カテゴリの傾向の発見を行っている。この研究では、ユーザーが、画像とテキスト情報が異なるツイートを行う場合が存在するため、テキスト情報を用いていない。まず初めに、2016 年の 1 月から 6 月までに収集した Twitter 画像を用いて CNN 特徴量の抽出を行い、クラスタリングを用いて画像のカテゴリ名をクラスターに割り当てている。クラスタリングには K-means を用いておりカテゴリ名をもとに、続いて、地域別に画像傾向、時期別傾向、時間別傾向の 3 つの分析を行っている。図 1 は地域別における画像傾向の結果であり、9 つの地域で 5 つの画像カテゴリの割合を示している。南米、アフリカでは人の画像が最も大きな割合を占めており、本研究で注目する食事画像カテゴリに関しては東アジアが最も大きな割合を示しており、続いて東南アジア、北米といった順となっている。この研究では、各地域における画像カテゴリの抽出にとどまっており、ひとつの画像カテゴリに注目したときの傾向については考慮されていない。そこで、本研究では食事画像に対して、さらに詳細なカテゴリ分類をし分析を行う他、同様にテキ

スト情報を用いずに行っていく。

Chang [2] らの研究では、地域特有の代表な服装の抽出とそれを用いた Fashion World Map を作成した。手法として、深層学習を用いて都市ごとに代表する全身を映したファッション画像の抽出を行う。抽出した画像から、都市を代表するアイテムを中間層の特徴マップと最適化手法を用いて各都市における代表的なアイテムを抽出している。画像は Web サイト Lookbook.nu(<https://lookbook.nu>) の画像を用いている。図 2 は各地域の代表画像が示されており、例として、東京では代表的な服装が黒色のジャケットであることが示されている。この研究では、各地域の代表的な服装を抽出することに重点を置いており、代表的な服装が複数あった場合にはより代表的であるほうの服装が抽出されるよう最適化が行われている。

そこで本研究では、これらの研究を参考に食事画像に注目して特徴量抽出による分析を行うことで、各地域における食事画像に対する地域的な特徴の発見を行う。長野ら [1] の研究では、5つの画像カテゴリーを用いていたが、本研究では、食事画像に限定することで食事画像のみの詳細な分析や、クラスタリングを行うことで、複数料理における地域的傾向の発見を行う。

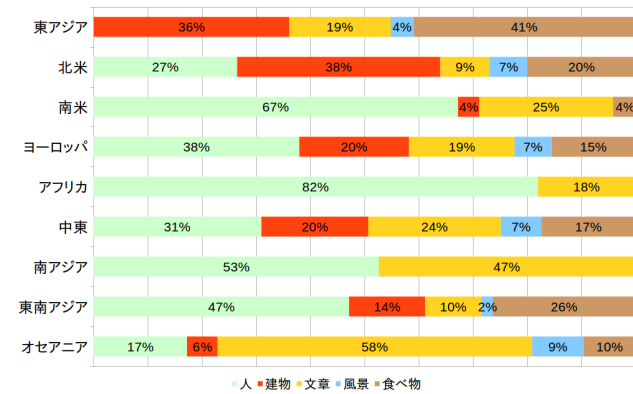


図 1 各地域ごとにおける画像カテゴリーの分布 (長野ら [1] から引用)



図 2 fashion map(Chang et al. [2] から引用)

2.2 CNN 特徴量を用いた研究

CNN 特徴量を用いた研究として、會下ら [3] の研究がある。この研究では、クラス分類学習と距離学習を用いたネットワークの CNN 特徴を用いることで、カロリー量推定精度の向上を行った研究である。距離学習は Triplet loss [4] を用いており、入力画像をクエリ画像、正画像、負画像の三枚組とし学習を行う。Triplet loss [4] は、クエリ画像と正画像のユークリッド距離が小さくなり、クエリ画像と負画像のユークリッド距離が大

きくなるように最適化を行っていく。モデル図は図 3 となっており、Triplet loss L_T の定義式は以下になっている。

$$L_T = \max(0, g + \|f_+ - f\|_2 - \|f_- - f\|_2) \quad (1)$$

f, f_+, f_- はそれぞれ、クエリ画像、正画像、負画像の CNN 特徴量である。定数 g は二種類の距離間のマージンであり、 $g = 0.3$ を用いており、本研究でもこの値を用いる。會下ら [3] の研究では、クラス分類学習と距離学習を用いることで画像検索精度が向上することを図 1 に示している。そこで、この研究を参考に、クラス分類学習と距離学習を行った CNN 特徴量を用いることで、より適切な地域特徴量の発見を試みる。

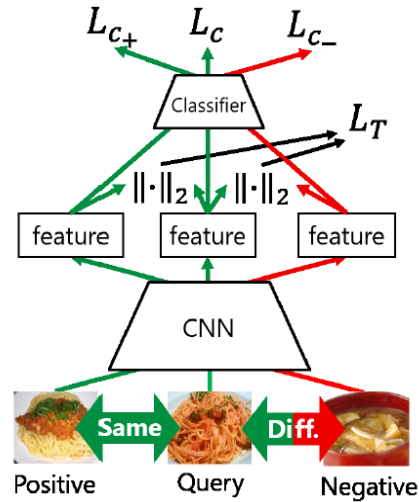


図 3 triplet loss の学習モデル (會下ら [3] から引用)

表 1 画像検索結果 (會下ら [3] から引用)

	mAP@1	mAP@5	mAP@10
VLAD (UEC. 100)	17.7	25.5	25.8
VGG16 (ImageNet) fc6	39.8	48.7	46.2
クラス分類学習 (UEC. 100)	56.7	65.4	62.3
クラス分類学習 (UEC. 256)	47.8	56.8	54.4
距離学習 (UEC. 100)	41.2	50.8	48.5
距離学習 (UEC. 256)	37.8	47.9	46.6
距離+クラス分類学習 (UEC. 100)	65.8	72.1	69.8
距離+クラス分類学習 (UEC. 256)	64.8	71.7	69.4

3. 手法概要

本研究では次の手順で、食事画像の地域特徴の分析を行った。図 4 は食事・非食事分類から地域分類までの流れ図である。

(1) 食事画像データセットを用いて、CNN をファインチューニングし食事・非食事分類器の作成。

(2) 作成した分類器を用いて Twitter 画像から食事画像を抽出。

(3) 画像データセットを用いて、Triplet loss で CNN をファインチューニングし食事特徴量抽出器の作成。

(4) 食事画像の特徴量の抽出。

- (5) 特徴量を用いた、クラスタリング。
- (6) クラスタリング結果を用いた各地域分析。

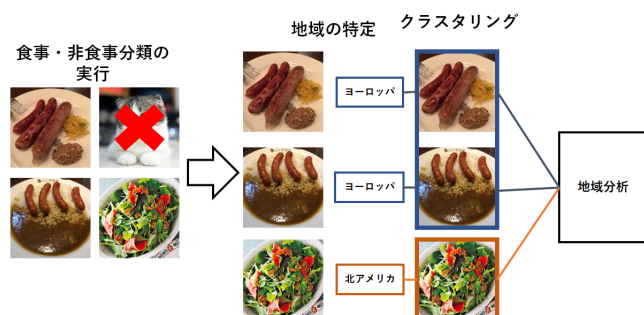


図4 食事・非食事分類から地域分析までの流れ図

4. 手法詳細

4.1 Twitter 画像における食事画像抽出

Twitter にはログデータが存在しており、このログデータから画像の投稿された時間や場所が取得できる。我々の研究室では、2011 年から現在に至るまでのログデータを保持しており、このログデータをもとに収集した 2016 年の 1 月から 12 月までの位置情報付き Twitter 画像を本研究では用いる。この Twitter 画像には、食事画像の他、人物、風景といったさまざまな種類の画像が含まれるため、食事画像のみを抽出する必要があり、そこで、食事・非食事分類器を作成した。食事・非食事分類器は、事前に食事・非食事画像を用いてネットワークを学習させることで、未知の画像に対しても高い精度で、食事か非食事を分類することができる。本研究では、食事・非食事分類器には ResNet [5] を使い、食事画像を UECFOOD100 [6] データセットと FOOD101 [7] データセットの画像、非食事画像には ImageNet [8] の 1000 種類分類タスクにもちられた画像と、河野ら [9] の研究で用いられた食事と誤認されやすい画像約 1 万 3 千枚を使用した。UECFOOD100 と FOOD101 はそれぞれ、約 1 万 4 千枚の画像と約 10 万枚の画像を含んでおり、これらの全ての画像と、ImageNet1000 はランダムに抽出した約 12 万枚の画像を使用する。UECFOOD100 [6] は東アジアの食事画像を、また FOOD101 [7] はスイスの食事画像を多く含むことから、2 つの食事画像データセットを食事・非食事分類器に学習させることで、多様な食事画像に対応できると考えられる。食事・非食事分類器を作成したのち、これを用いて Twitter 画像から食事画像の抽出を行う。

4.2 食事特徴量の抽出

抽出された食事画像から地域分析を行うために、CNN 特徴量を用いた食事特徴量を抽出する。CNN 特徴量は、一般的に、事前に学習したネットワークの中間層の活性化値が用いられ、従来の特徴量よりも画像の意味を考慮した特徴量を得ることができる。地域傾向分析に用いる食事特徴量は、3 つのそれぞれ学習させたネットワークから抽出した食事特徴量を用いてクラスタリングを行い、どの食事特徴量がより地域傾向を分析しやすいかを調べることで決定する。食事特徴量の抽出に使用するネットワークは、VGG16 [10] に対してそれぞれ ImageNet [8] の 1000 種類分類タスクに用いられた画像を学習させた ImageNet

CNN、食事画像を学習させた Food CNN、距離学習を用いて食事画像学習させた Triplet Food CNN とする。ImageNet には、Python の機械学習用フレームワークである PyTorch の学習済みモデルを用いる。Food CNN と Triplet Food CNN の学習には、UECFOOD100 [6] と FOOD101 [7] の画像の中からあらかじめ分類しやすいと思われる 106 種類の食事画像を用いる。Food CNN は、VGG16 ネットワークの最終層の出力を 106 クラスとしてファインチューニングを行い食事画像を学習させたネットワークを用いる。Triplet Food CNN は、最終層を Food CNN と同じ 106 クラスとして距離学習とクラス分類学習を同時に行ったネットワークを用いる。距離学習に使用する食事画像は、クエリ画像に対してランダムに選択した同じクラスの画像を正画像、異なるクラスの画像を負画像とした 3 枚 1 組の画像とした。それぞれ学習を行った特徴量抽出器から食事特徴量を抽出、その後、食事特徴量のクラスタリングを行い、どれほどクラスタ内の画像間の非類似度を調べることで、分析に用いる食事特徴量を決定する。特徴量は VGG16 [10] の fc7 の 4096 次元のベクトルを L2 ノルムで正規化し、長野らの研究 [1] を参考に、Python のオープンソースの scikit-learn のライブラリにある PCA を用いて次元圧縮を行い、128 次元の特徴量を用いる。

4.3 食事画像のクラスタリングと画像の比較

抽出された食事画像の特徴量を分析するため、クラスタリングを行う。クラスタリングを行うことで、食事画像をある程度まとまった料理クラスターに分けることができる。まず初めに、各特徴量抽出器から抽出された食事特徴量の中から、地域傾向分析に用いる食事特徴量を決定するため、特徴量抽出器ごとの平均画像非類似度とその分散を求める。クラスタ内に含まれる画像間の非類似度をユークリッド距離を用いて求めたのち、クラスタ内の画像非類似度の平均値とその分散値を求める。ユークリッド距離を用いた非類似度は、特徴量を L2 ノルムで正規化していることから、0 に近いほど類似した画像同士であり同一の画像で距離を求めた場合は値がほぼ 0 となる。また、分散が低いほど、特徴量にばらつきがなくクラスター内に類似した画像多く含まれており、分散が大きいほど、類似しない画像が多く含まれる。まとめると、平均画像非類似度と分散値がともに小さいほど、よりまとまったクラスターと言える。そこで本研究では、クラスタ内の平均非類似度と分散値の平均をとることで、各食事特徴量抽出器の平均非類似度と平均分散値とし、この 2 つの値を見ることで、分析に用いる食事特徴量を決定する。次に、食事特徴量を用いたクラスタリング結果から、作成されたクラスターに料理カテゴリを手作業でつける。その後、料理カテゴリごとに地域割合を調べることで各地域における地域傾向の発見を試みる。クラスタリングに用いた手法として、K-means を使用する。K-means は、あらかじめクラスター数を決め、与えられたベクトルに対するクラスター割り当てとクラスター中心の計算を繰り返すことでクラスタリングを行う手法で、クラスタリングにおいて一般的に使用されている手法である。本実験ではクラスター数を 100、150、200 に設定し、各特徴量抽出器ごとにクラスタリングを行う。

4.4 VisualRank

クラスタリングから算出されたクラスタに対して、料理カテゴリを割り当てる際に、類似しない画像が含まれており、適当な料理カテゴリを割り当てるのが難しい場合がある。そこで、本研究では VisualRank [11] を用いてクラスタ内の画像のランク付けを行う。VisualRank [11] は画像間の類似度をもとに画像のランク付けを行う手法である。各画像の VisualRank 値からなるランキングベクトル R を、繰り返し更新させ、収束させることで得られる。 S は、クラスタの画像数 $\times$ クラスタの画像数 次元の画像間距離行列であり、要素の値が大きいほど、画像の類似度が高くなる。ユークリッド距離を用いた非類似度は小さいほど類似することから、クラスタ内の画像非類似のうち最大値から、各距離を引き類似度として、 S に用い、各列の合計が1になるように正規化を行い VisualRank 値が変化しないようにして用いる。 P は補正ベクトルとされ、今回はクラスタ内の画像数の逆数を要素とする列ベクトルを用いる。 α は補正值であり、本研究では、 $\alpha = 0.8$ とする。クラスタ内の画像にランク付けを行うことで、クラスタ内に含まれる類似画像を確認しやすく、容易に料理カテゴリを割り振ることができる。

$$R = \alpha(S \times R) + (1 - \alpha)P, (0 \leq \alpha \leq 1) \quad (2)$$

4.5 地域分析

クラスタリング結果から料理名を割り振ったクラスタをもとに、画像の位置情報をもとに分析を行う。まず初めに、料理ごとの食事画像における地域傾向を分析する。クラスタリングによって作成したクラスタは、料理ごとにある程度のまとまりのある画像を含んでいる。そこで、クラスタに料理カテゴリーを割り当て、料理ごとにクラスタ内の画像の位置情報を集計する。その後、各料理における地域ごとの傾向を求めることで、料理ごとの食事画像の地域傾向を分析する。また、地域ごとにクラスタごとの割合も求めることで、地域ごとの食事画像傾向を分析する。

4.6 Grad-CAM

最後に食事画像の特徴箇所の可視化を行い、どの部分が地域特徴を示しているかを調べる。本研究では、可視化に用いる手法として、Grad-CAM [12] を用いる。Grad-CAM は、ネットワークの最終層で出力されたクラス c におけるスコア y^c に対して、最終畳み込み層の特徴量マップ A_{ij}^k について微分した値である勾配 ∂A_{ij}^k を、チャンネル k ごとに全ピクセル i, j において平均をとり重み α_k^c を求める。その後、式 4 により、チャンネルにおける $\alpha_k^c A^k$ の合計値に対して関数 $ReLU(x) = \max(x, 0)$ をかけることで、マスク画像 L^c が生成し、元画像と掛け合わせることで、画像のどの位置が活性化しているかを見ることができる。Grad-CAM [12] に用いるネットワークは、出力を各地域クラスとしたネットワークに各地域から、(ある料理、または全料理)の食事画像を学習させたネットワークを用い、食事画像のより詳細な地域分析を行う。

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (3)$$

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (4)$$

5. 実験

5.1 食事画像

まず初めに、作成した食事・非食事分類器を用いて、画像分類を行う。ログデータをもとに収集した、2016 年 1 月から 12 月までの 1 年間の位置情報付きツイートに含まれていた画像約 378 万枚を分類器を用いて食事画像の抽出を行う。約 22 万枚の画像が抽出された後、そこから分析を正確にするため画像のカラーヒストグラムをもとに、重複する画像を取り除いた。最終的に約 19 万枚の食事画像が抽出され、これらの画像を用いて特徴量抽出を行う。各地域ごとの全画像の枚数は表 2、各地域ごとの画像枚数は表 3 の通りである。

表 2 各地域ごとの全画像の枚数

地域	画像枚数 (枚)
東アジア	488,609
南アジア	79,179
東南アジア	609,671
北アメリカ	786,093
南アメリカ	410,086
ヨーロッパ	791,095
アフリカ	150,550
オセアニア	34,543
中東	434,790
合計	3,784,616

表 3 各地域ごとの食事画像の枚数

地域	画像枚数 (枚)
東アジア	69,826
南アジア	999
東南アジア	59,459
北アメリカ	23,867
南アメリカ	8,211
ヨーロッパ	14,572
アフリカ	1,229
オセアニア	914
中東	14,115
合計	193,192

5.2 特徴量抽出と比較

抽出された食事画像に対して、ImageNet CNN、Food CNN、Triplet Food CNN の 3 種類のネットワークを用いて特徴量抽出を行い、K-means を用いてクラスタリングを行った。クラスタリングを行う際に、全食事画像に対してクラスタリングを行った場合と、食事画像を各地域に分けてクラスタリングを行った場合の平均画像非類似値と平均分散値を求め、比較を行った。平均画像類似値は0に近いほど、よりクラスタ内に類似した画像を含んでおり、平均分散値が低いほど、クラスタ内に類似した画像が含まれている。評価は、平均画像非類似度と平均分散値を考慮し行った。

クラスタリングを行った際に、全食事画像を用いてクラスタリングを行った場合よりも、食事画像を各地域ごとに分けてクラスタリングを行ったほうが、平均分散値が低いという結果となった。食事画像を各地域ごとに分けてクラスタリングを行ったほうが類似した画像の割合が多いクラスタを生成することが分かった。このクラスタ内に含まれる画像は、全食事画像を用いたクラスタリングでは、別々のクラスタおよびそれぞれ食事カテゴリが異なるクラスタに割り当てられた。

各特徴量ごとに、結果を見た際に、平均分散値が最も低い特徴量は ImageNet CNN を用いた場合となった。しかし、ImageNet CNN を用いた特徴量の場合、平均非類似度が高い結果となり、クラスタ内の画像が類似しているのかは信憑性に欠ける。また、Food CNN を用いた特徴量は平均分散値が最も高い値を示しており、クラスタ内の類似画像にばらつきがある。そ

表 4 各特徴量ごとの平均類似度と平均分散値

クラスタリング	全画像		各地域ごと	
	平均画像非類似度	平均分散値	平均画像非類似度	平均分散値
ImageNet CNN(クラスタ数 100)	1.1366	0.0139	1.1186	0.0141
Food CNN (クラスタ数 100)	1.0898	0.0172	1.0828	0.0165
Triplet Food CNN (クラスタ数 100)	1.0931	0.0171	1.0888	0.0163
ImageNet CNN(クラスタ数 150)	1.1193	0.0135	1.0980	0.0133
Food CNN (クラスタ数 150)	1.0715	0.0165	1.0569	0.0158
Triplet Food CNN (クラスタ数 150)	1.0705	0.0161	1.0673	0.0154
ImageNet CNN(クラスタ数 200)	1.1081	0.0131	1.0814	0.0127
Food CNN (クラスタ数 200)	1.0564	0.0161	1.0387	0.0149
Triplet Food CNN (クラスタ数 200)	1.0568	0.0156	1.0496	0.0145

ここで、分析には 2 番目に平均分散値が低く、會下ら [3] の研究から画像検索精度が高かった、Triplet Food CNN の特徴量を用い、食事画像を各地域ごとに分けて行ったクラスタリング結果に対して分析を行った。クラスタ数は、食事画像を分割してクラスタリングを行うため、クラスタ内の画像数をある程度保つ必要があることから 100 とした結果を分析に使用した。

5.3 クラスタリング結果の分析

Triplet Fodd CNN から抽出した特徴量から、K-means による地域別のクラスタリングをし、分析を行った。K-means のクラスター数は 100 を用いた。クラスタリングによって求められたクラスタに対して、料理カテゴリを割り振る際に、クラスタ内の画像に対して VisualRank [11] を求めた。これにより、VisualRank 値の大きい画像を見ることで、より適切な料理カテゴリを割り振ることができる。また、分析の際に VisualRank 値が低い画像は、分析結果に対して影響を与える可能性があるため、画像の VisualRank 値にクラスタに含む画像枚数の値をかけたものに対して、閾値を設け、その値よりも大きい画像を分析に用いた。閾値は、各地域ごとの画像の減り方を考慮し 0.821 とした。クラスタ内の画像が、散々しており料理カテゴリが特定できないクラスタは分析を行わなかった。また、分析を行う際に、南アジア、アフリカとオセアニアの画像はクラスタリングを行った際にクラスタ内の画像数が分析に使用できるほど確保できなかったため、これら 3 地域を除いた、東アジア、東南アジア、北アメリカ、南アメリカ、ヨーロッパ、オセアニアの 6 地域で分析を行った。分析を行った地域は図 6 で色を付けた箇所である。クラスタに割り当てた料理カテゴリは表 5 の 17 カテゴリである。当初、料理カテゴリは飲料、デザート、パン類、ご飯類、肉料理、麺類、魚貝類、サラダ、スープ類の 9 種類に定め分析を行った。しかし、クラスタリング画像を観察していく際に、ポテトやハンバーガーといったファストフードと思われるクラスタが多数できることを確認し、ファストフードを料理カテゴリに追加した。カレーに関して東アジア以外の地域でご飯が含まれていない画像が見られたため、ご飯類と分離した。また、既存の料理カテゴリでは分類できないものを追加していった他、中華、チーズに関しては当初予定したの料理カテゴリで分類できないが、クラスタ内に料理カテゴリを割り当てれるほど、画像がまとまっていたために料理カテゴリに追加した。中華とチーズに関しては、麻婆豆腐や酢豚といった食事画像に対して中華カテゴリ、ライスやパスタにチーズを和え

た食事画像に対してチーズカテゴリと分類を行った。結果として 17 種類の料理カテゴリを分析に用いた。

表 5 食事カテゴリの一覧

食事カテゴリ		
飲料	デザート	パン類
ご飯類	揚げ物	炒め物
焼き物	中華	カレー
肉料理	麺類	卵料理
サラダ	魚貝類	スープ
チーズ	ファストフード	

5.3.1 地域別傾向分析

まず初めに、地域別にどの料理カテゴリが多いかを分析した。分析結果は図 5 に示しており、上位 5 位までの料理カテゴリの割合を太字で表している。東アジアの結果では、麺類の食事画像が最も多く、29.3%という結果となった。麺類の中でも、ラーメンの画像が非常に多く 14 つのクラスタ内でラーメンの画像が見られた。2 番目に多い画像はご飯類の画像で 14.5%という結果となった。東南アジアの結果では、肉料理が最も多く 16.6%続いてデザートが 14.8%という結果となった。北アメリカの結果では、アジアの食事傾向とは違い、パン類が 18.6%、続いてサラダが 15.2%という結果になった。南アメリカの結果では、デザートが最も多く 25.1%、続いてパン類が 22.5%という結果になった。ヨーロッパの結果として、北アメリカと南アメリカと上位 4 項目が等しく、また、最も多いカテゴリはパン類で 18.7%、次に多いのがデザートで 25.1%となった。中東では、ヨーロッパと上位 5 項目が同じ項目となり、最も多いカテゴリは飲料で 23.2%、次に多いのがデザートで 22.0%という結果となった。

5.3.2 料理別における地域分布の分析

次に、料理カテゴリごとにおける地域分析を行った。料理カテゴリごとに各地域の画像枚数を計算し、その料理の合計枚数を表 7 に、その料理における各地域の割合を表 6 にまとめた。料理カテゴリのなかで最も多い画像は肉料理となり、続いて麺類、デザート順となった。また、料理カテゴリごとに地域ごとの割合を求めた。その中で、全画像における地域内の画像割合よりも 10%以上高いものに関して注目した。東アジアにおいて、最も高い割合を示している料理カテゴリは、麺類、ご飯類、揚げ物、焼き物、魚貝類、卵料理、カレー、中華となって

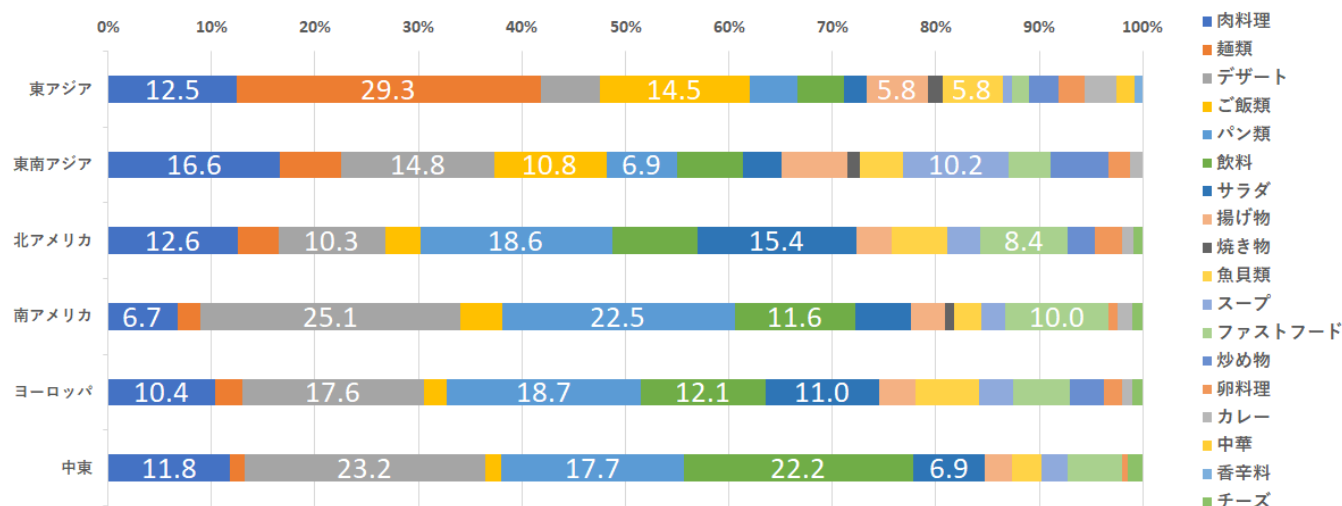


図 5 地域における料理カテゴリ割合（白文字は、上位 5 項目の割合を示す）

表 6 食事における地域傾向（太文字は全料理よりも 10%以上高い割合を示す）

料理カテゴリ	東アジア (%)	東南アジア (%)	北アメリカ (%)	南アメリカ (%)	ヨーロッパ (%)	中東 (%)
全料理	38.77	30.71	12.52	3.72	7.41	6.85
肉料理	36.32	38.34	11.65	1.87	5.76	6.03
麺類	81.00	12.90	3.44	0.60	1.37	0.70
デザート	18.68	38.37	10.75	7.87	11.00	13.34
ご飯類	57.49	33.98	4.25	1.53	1.71	1.05
パン類	18.60	21.92	23.85	8.71	14.41	12.50
飲料	22.92	26.02	13.50	5.72	11.86	19.97
サラダ	16.21	21.15	35.10	3.72	15.07	8.74
揚げ物	43.50	37.66	8.13	2.32	5.00	3.39
焼き物	43.49	37.66	8.13	2.32	5.00	3.39
魚貝類	45.13	26.45	13.29	1.95	9.25	3.92
スープ	14.67	66.26	8.39	1.82	5.23	3.63
ファストフード	15.17	30.92	25.71	9.23	10.09	8.87
炒め物	32.98	50.30	9.71	0.00	7.00	0.00
卵料理	46.34	29.25	15.0	1.49	6.10	1.79
カレー	64.19	21.33	7.56	2.79	4.12	0.00
中華	100.00	0.00	0.00	0.00	0.00	0.00
チーズ	0.00	0.00	33.72	12.03	23.08	31.16

おり、そのなかで全画像の割合よりも 10%以上高いものは、麺類、ご飯類、カレー、中華となった。東南アジアにおいて最も高い割合を示しているものは料理カテゴリは、肉料理、デザート、飲料、スープ、炒め物となっており、そのなかで全画像の割合よりも 10%以上高い割合を示しているものは、スープと炒め物となった。北アメリカにおいて最も高い割合を示しているものは、パン類、サラダ類、チーズとなっており、その中で、全画像の割合よりも 10%以上高い割合を示しているものは、パン類、サラダ、ファストフード、チーズとなった。南アメリカ、ヨーロッパでは、最も高い料理カテゴリは存在せず、また全画像の割合よりも 10%以上高い割合を示しているものも、存在しなかった。中東で、最も高い料理カテゴリは存在しないが、全画像の割合よりも 10%以上高い割合を示しているものは、飲料となった。

5.3.3 特徴量の可視化

最後に食事画像のどの部分に地域特徴量が表れているのかを

調べるため、Grad-CAM [12] における可視化を行った。出力層を東アジア、東南アジア、北アメリカ、南アメリカ、ヨーロッパ、中東の 6 地域とした VGG16 [10] ネットワークに対して、それぞれの料理カテゴリの画像を学習させ、そのモデルを用いて Grad-CAM [12] を用いて可視化をさせた。本研究では、肉料理とパン類に絞り小規模に実験を行った。図 7 と図 8 は実際に地域特徴箇所が表示された例である。それぞれの画像は、地域別のクラスタリング結果においてその地域にしか現れなかったクラスタ内の画像である。図 7 は、東南アジアにおける肉類の画像である。丸い部分が特徴的であり、Grad-CAM [12] を用いた結果、肉の丸い部分に地域特徴が表れているのが分かる。図 8 は、中東におけるデザートの画像である。バクラバアといわれるデザートで、中東において人気のある食べ物であり、実験の結果パイ生地の色緑色の部分に地域特徴が表れた。

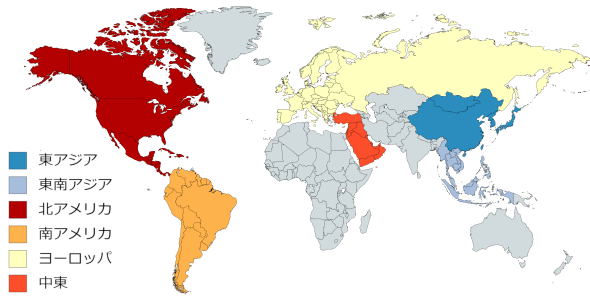


図 6 分析を行った地域マップ

表 7 各料理における画像枚数

料理カテゴリ	画像枚数 (枚)	料理カテゴリ	画像枚数 (枚)
肉料理	20,998	魚貝類	7,797
麺類	22,143	スープ	7,445
デザート	18,682	ファストフード	6,355
ご飯類	15,404	炒め物	5,330
パン類	15,165	卵料理	3,413
飲料	11,902	カレー	2,935
サラダ	8,526	中華	1,017
揚げ物	8,173	チーズ	507
焼き物	8,173		

6. 考 察

6.1 地域別傾向分析について

5.3.1 項の実験では、地域別にどの料理カテゴリが多いかを調べた。図 9 から図 13 までの図は各地域における特徴的な食事画像の一例である。東アジアの上位 5 項目を見た際に、麺類、ご飯類、肉料理、揚げ物、魚貝類といった他地域と比べあまり見られない料理カテゴリが挙げられている。麺類が 29.3%という高い割合になった理由として、図 9 としたラーメンの画像非常に多いことが挙げられ、東アジアのクラスターのうち 14 クラスターでラーメン画像が見られた。また、ラーメン以外にもうどん、そば、パスタといった様々な画像が多くみられたため、東アジアを代表とする料理カテゴリであるといえる。次に、ご飯類の割合が高い理由として、アジアの主食であることから、定食といった複数食材に対してもご飯類の画像が含まれておりご飯類の割合が高いものとなったと考えられる。これは、東南アジアにも同じことが言える。また、上位 5 項目を他の地域と比較したときに、東アジアの傾向は他の地域と類似しておらず独特な傾向を示していることが分かった。東南アジアの食事傾向において最も注目したい点は、スープが上位に入っていることである。このスープの多くは、図 10 のような汁に野菜や肉を煮た料理が多く、東南アジアの特徴的な料理カテゴリであるといえる。また、肉料理において、他地域では見られない、丸い肉にポテトサラダとった図 11 のような料理も見られた。それ以外の部分に関しては、他の地域と類似した食事傾向が表れている。北アメリカの傾向とアジアの傾向を比較した際に、パン類とサラダのカテゴリが多い傾向となった。北アメリカのパン類の料理には、ピザが最も多く含まれている他、サラダはグリーンサラダやトマト、チーズを和えたものが多くなった。東アジアでは、図 12 のようなサラダしか見られず、北アメリカのサ

ラダは東アジアよりも多様性に富んだ料理が含まれていることが分かった。一方、南アメリカ、ヨーロッパと比較した際には、料理の上位 5 項目のうち 4 項目が同じカテゴリであり、類似した食事傾向をもつことが分かった。また、北アメリカ、南アメリカの両方にファストフードが上位 5 項目に入っており、他地域よりもアメリカにおいてファストフードが人気を誇っていることが伺える他、南アメリカにおいては、北アメリカと比べた際にデザートがチョコレートを用いたデザートがより多く見受けられた。中東においては、料理カテゴリ 5 項目がヨーロッパと等しくなったが、この地域では、図 13 のような茶色のコーヒーマーといった料理が多くみられ、ヨーロッパと類似しながらも、独特な地域傾向を持つことが分かった。

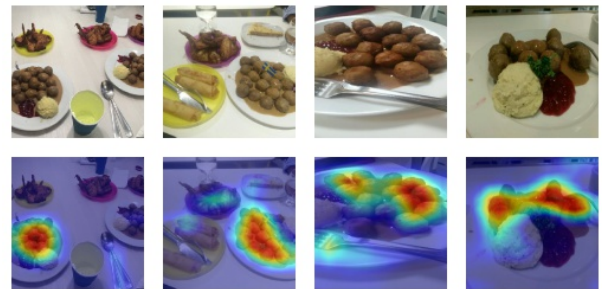


図 7 東南アジアにおける肉類の画像に対する Grad-CAM の適用例

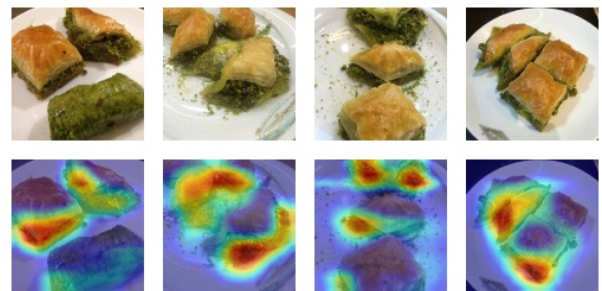


図 8 中東のデザートの画像に対する Grad-CAM の適用例

6.2 料理別における地域分布について

5.3.2 項の実験では、料理別にどの地域の割合を調べており、全体の画像割合から 10%以上を上昇したのに関して注目した。まず初めに、麺類であるが、東アジアにおける割合が最も高くなり 81.0%となった。多くのラーメンが食されている他、ラーメン以外の麺類も食されていることが要因であると考えられる。ご飯類も東アジアの割合が最も高い値となる。麺類と比較した際に、ご飯類は東南アジアでも 34.0%あるが、麺類に関しては、東南アジアの割合が 12.8%と小さくなることが分かる。東南アジアではスープの一部が麺類として食べられており、麺類の割合が低くなっているからである。また、ご飯類とカレーを比較すると、東アジアの割合が最も高いといった類似の傾向が得られる。東アジアや東南アジアではご飯類とカレーは一緒に食べられることが多い料理であることが要因として考えられる。スープに関しては、東南アジアが最も割合が高く、66.3%となった。これは、東南アジアのスープは野菜、肉や麺類といった他のカテゴリの食材を多く使った画像が見られ、他の地域と比べて幅広く多様性に含んだ料理となったと考えられる。中華に関しては、東南アジアが 100%となった。これは、こ

の中華料理そのものが東南アジア独特の料理であり、地域特有の料理であることが分かる。パン類とサラダ類は、北アメリカにおける割合が最も高くなり 23.9%と 35.1%になっており、東アジアと東南アジアよりも高い割合となった。全画像に対しての割合と比較する際に、ヨーロッパではパン類、サラダ類がともに増加傾向にあるのに対して、南アメリカでは、パン類のみ増加傾向を示すことが分かった。チーズに関して言えば、北アメリカが最も高くなり、33.7%となっている。また、注目すべき点としては、画像枚数の多い東アジア、東南アジアに見られていない点であり、北アメリカやヨーロッパといった欧米付近で人気のある料理といえる。飲料において、東南アジアが最も高い割合となった。しかし中東の割合を見てみると 20.0%と全画像に対する割合よりも大きい結果となり、少なからず飲料においての地域分布に影響を与えているのが分かる。ファストフードにおいては、東南アジアが 30.9%と最も多い割合を占めているが、北アメリカも 25.7%と競争力のあることが示している。

6.3 可視化について

5.3.3 章の実験では、料理における地域特徴量箇所の可視化を行った。実験に使った画像に対して顕著に地域特徴量が表れる結果となった。これは、Grad-CAM に用いた画像がその地域にしか存在しない画像であったからである。今回行った分析は、小規模にとどまっており、2つの地域間における分析を行うことでより食事画像のより詳細な部分の分析が可能だと考えられる。

7. ま と め

今回、テキスト情報を使わず、位置情報付き食事画像のみをもとに CNN 特徴量を用いたクラスタリング手法により、食事画像における地域傾向の発見を行うことができた。CNN 特徴に関しては Triplet loss [4] を用いており、クラスタリングに K-means を用いた。また、Grad-CAM [12] を用いた可視化手法により地域特徴を表している部分の発見を行うこともできた。分析の結果から、地域別に独自の食事画像傾向や他の地域と類似した食事画像傾向を見られたほか、食事画像ごとに割合を占める地域を確認できたことや、食事の形、食材といった部分に地域特徴をみることができた。

8. 今後の課題

今回、分析を行う地域数を絞って行ったため、南アジア、アフリカ、オセアニアの3地域の地域傾向を発見することが出来なかった。これは、3つの地域において食事画像が極端に少ないことに起因しており、画像枚数の偏りをなくすためこれら3地域の位置情報付き画像を集める必要がある。しかし、長野ら [1] 研究から、これらの3地域において Twitter に投稿されている食事画像が極めて少ないことが分かっており、課題の一つである。また、より詳細な分析を行うためには、料理カテゴリを細分化する必要がある。しかし、カテゴリ数が多いほど、データの精査に時間と手間がかかる他、より表現力のある CNN 特徴量が求められるため、データ精査の効率化と、より分析に適切な CNN 特徴が必要となる。



図 9 東アジアにおける麺類の画像例



図 10 東南アジアにおけるスープの画像例

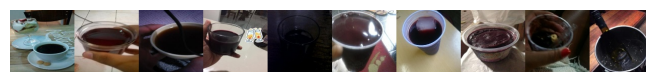


図 11 東南アジアにおける地域特有の肉料理の画像例



図 12 北アメリカにおけるサラダの画像例



図 13 中東における地域特有の飲料の画像例

文 献

- [1] 長野哲也, 柳井啓司. Twitter 画像の大規模時空間分析. データ工学と情報マネジメントに関するフォーラム (DEIM), 2017.
- [2] Y. T. Chang, W. H. Cheng, B. Wu, and K. L. Hua. Fashion World Map: Understanding Cities Through Streetwear. In *Proc. of ACM International Conference Multimedia*, pp. 91–99, 2017.
- [3] 會下拓実, 下田和, 柳井啓司. CNN 特徴量学習に基づく画像検索による食事画像カロリー量推定. 電子情報通信学会論文誌 D, No. 8, pp. 1099–1109, 2018.
- [4] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, B. Philbin, J. Chen, and Y. Wu. Learning finegrained image similarity with deep ranking. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2014.
- [5] K. He, X. Zhang, and J. Ren, S. and Sun. Deep Residual Learning for Image Recognition. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2015.
- [6] Y. Matsuda, H. Hoashi, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In *Proc. of IEEE International Conference on Multimedia and Expo (ICME)*, 2012.
- [7] B. Lukas, G. Matthieu, and V.G. Luc. Food-101 – mining discriminative components with random forests. In *Proc. of European Conference on Computer Vision*, 2014.
- [8] A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2012.
- [9] 河野憲之, 柳井啓司. ラーメン vs カレー: 2年分のログデータと高速食事画像認識エンジンを用いた twitter 食事画像分析とデータセット自動構築. 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU), 2013.
- [10] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, Vol. abs/1409.1556, , 2015.
- [11] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 11, pp. 1870–1890, 2008.
- [12] R.R. Selvaraju, A. Cogswell, M. and Das, R. Vedantam, D. Parikh, and D. Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2016.