

大量の Twitter 位置情報付き画像を用いた 世界各地域における食事傾向分析

岡本 開夢^{1,a)} 柳井 啓司^{b)}

概要

Twitter は若者を中心に世界中で普及しており、ツイートとともに多くの画像が投稿されている。それらの画像のなかでも食事画像は地域傾向があるとされている。そこで本研究では、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行う。分析手法としては、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、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。

1. はじめに

Twitter は若者を中心に世界中で普及しており、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、地域的傾向分析といった多くの分析に使用されている。長野ら [6] [12] の研究から、各地域における Twitter 画像のカテゴリ割合が知られている。また、その中でも食事画像に注目した際に地域的傾向があると予想されている。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。分析を行う際には、テキスト情報を使わずに、CNN 特徴量とクラスタリングのみを用いて行う。これは、画像付きツイートのテキスト情報が、必ずしも適切に画像を表現しているとは限らず、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場合があるからである。また、食事画像においてどの部分が地域特有な箇所なのかを明らかにするため、Grad-CAM を用いて可視化を試みる。

2. 関連研究

2.1 地域分析

地域分析を行った先行研究として、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 [6] の研究がある。長野ら [6] 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、まず 2016 年の 1 月から 6 月までに収集した Twitter 画像を用いて CNN 特徴量の抽出を行っている。次に、CNN 特徴量をもとに K-means によるクラスタリングを行うことで、画像クラスタを作成し、作成したクラスタの画像カテゴリから地域傾向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。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のみに注目し、さらに詳細な地域分析を行っていく。

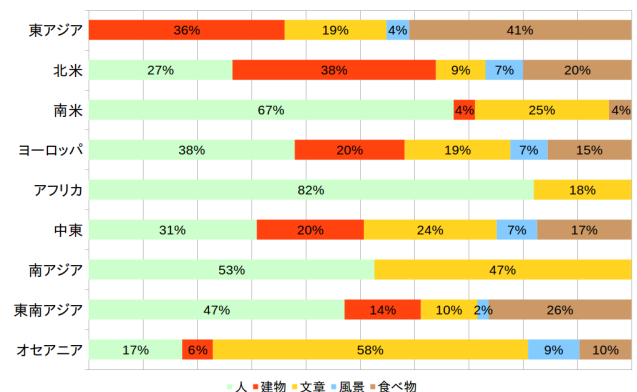


図 1 各地域ごとにおける画像カテゴリの分布 (長野ら [12] から引用)

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 [13] 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 特徴量を用いることで、カロリー量推定精度の向上を行った研究である。距離学習には Triplet loss [9] 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 [9] 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。Triplet loss L_T の定義式は以下のようになっている。

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

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。結果の表 1 から画像検索精度が向上することを示しており、本研究でも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。

¹ 電気通信大学院 情報理工学研究科 情報学専攻

^{a)} okamoto-ka@mm.inf.uec.ac.jp

^{b)} yanai@cs.uec.ac.jp

表 1 画像検索結果 (會下ら [13] 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 2 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・非食事分類器の作成、食事画像の分類
- (2) Triplet loss を用いた食事特徴量抽出器の作成、食事画像の特徴量抽出
- (3) 食事特徴量を用いた、クラスタリングと各地域分析
- (4) 食事画像における地域特有箇所の可視化



図 2 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分類

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 画像から食事画像のみを抽出するため、今回、食事・非食事分類器の作成を行う。食事・非食事分類器には ResNet[1] を用い、食事画像を UECFOOD100 [5] データセットと FOOD101 [4] データセットの画像、非食事画像には ImageNet [3] の 1000 種類分類タスクにもちられた画像、河野ら [10] 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 千枚を学習させ作成する。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、ImageNet からランダムに抽出した約 12 万枚の画像を使用する。

4.2 食事特徴量の抽出

抽出された食事画像の特徴量をもとにクラスタリングする際に、本研究では CNN 特徴量を用いた食事特徴量を抽

出する。地域傾向分析に用いる食事特徴量は、會下らの研究と同様にクラス分類学習と距離学習を用いた VGG16 [8] の CNN 特徴量を用いる。ネットワークの学習には、Triplet loss [9] を用いて、UECFOOD100 [5] と FOOD101 [4] の画像の中からあらかじめ分類しやすいと思われる 106 種類の画像で学習を行う。Triplet loss 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、VGG16 [8] の最終層の出力層を 106 クラスとしたものを用いて学習させる。CNN 特徴量はネットワークの fc7 層の 4096 次元のベクトルを L2 ノルムで正規化したのち、PCA を用いて次元圧縮を行い、128 次元まで落とし込み用いる。

4.3 食事画像のクラスタリングと地域分析

次に、分類された食事画像の特徴量を用いて、位置情報から算出した画像の地域ごとに K-means を用いたクラスタリングを行う。クラスタ数を 100 として、作成されたクラスタに料理カテゴリを手作業で割り振り、分析を行う。地域ごとに料理カテゴリの割合を調べることで地域における食事傾向分析と、料理カテゴリごとに地域の割合を調べることで食事における地域分析の 2 項目について行う。

4.4 VisualRank

本研究では画像間の類似度をもとに川久保らの研究 [11] でも用いられた VisualRank [2] を使用してクラスタ内の画像のランク付けを行う。この手法を用いることでクラスタ内に多く含む画像が高い VisualRank 値を持つため、そのクラスタを代表する画像群を取得することができる。これにより、クラスタ内に類似しない画像が含む場合でもより適当な料理カテゴリを割り振ることができる。VisualRank [2] は各画像の 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、要素の値が大きい画像ほど、クラスタ内に類似する画像が多くなる。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、 $\alpha = 0.8$ として用いた。

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

4.5 Grad-CAM による可視化

最後に食事画像のどの部分が地域特徴を示しているかを調べるため Grad-CAM [7] を用いて可視化を行う。可視化を行うことで、例えば、日本におけるうなぎであればうなぎの蒲焼の部分といった、料理におけるその地域特有な形状や調理法の発見を試みる。Grad-CAM は、画像のどの部分が分類クラスを決定する際に影響を与えているかを、特定する手法であり、式 3 と 4 によって求められる。式 3 により、重み係数 α_k^c を、ネットワークの最終層で出力されるクラス 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 において平均をとることで求める。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、画像のどの部分が活性しているかを見ることができる。

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

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

5. 実験

5.1 食事・非食事画像分類

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、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分類を行う。約 22 万枚の画像が分類された後、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 万枚の食事画像が分類され、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と各地域ごとの画像枚数は表 2 の通りである。また、クラスタリングを行う際に画像枚数が少なくクラスタに対して料理カテゴリを割り振るのが難しく分析ができない、南アジア、アフリカとオセアニアの 3 地域においては分析を行わず、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、中東の 6 地域で分析を行った。

表 2 地域における画像枚数

画像枚数	東アジア	南アジア	東南アジア	北アメリカ	南アメリカ
全画像	488,609	79,179	609,671	786,093	410,086
食事画像	69,826	999	59,459	23,867	8,211
画像枚数	ヨーロッパ	アフリカ	オセアニア	中東	合計
全画像	791,095	150,550	34,543	434,790	3,784,616
食事画像	14,572	1,229	914	14,115	193,192

5.2 クラスタリングと地域分析

食事画像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。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 [2] を求めた。VisualRank 値の高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値に対して、閾値を設けその値よりも小さい画像を分析から除いた。閾値は各地域ごとの画像の減り方を考慮し 0.821 とし、画像が散々しているクラスタに関しては、料理カテゴリが特定できないため分析から除いた。クラスタに割り当てた料理カテゴリは表 3 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、クラスタ画像を観察していく際に、この 9 種類では分類できないが、画像がまとまっているカテゴリが多数みられることから料理カテゴリを追加し、結果として 17 種類の料理カテゴリを分析に用いた。

5.2.1 地域別食事傾向分析

まず初めに、地域別に料理カテゴリの割合を求め分析を行った。分析結果は図 3 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おり、図 6 から図 8 までの図は各地域における食事画像の一例である。東アジアの

表 3 食事カテゴリの一覧

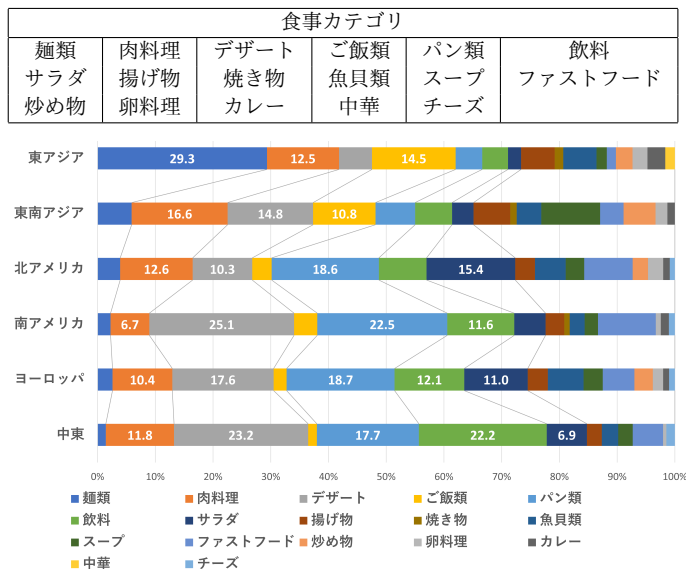


図 3 地域別における料理カテゴリ割合（白文字は、上位 5 項目の割合を示す）

食事傾向として、麺類、ご飯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上位に含まれている。麺類が 29.3% という高い割合になった理由として、図 6 といったラーメンの画像非常に多く見られたほか、ラーメン以外にもうどん、そばといった様々な画像が多くみられ東アジアを代表するカテゴリであるからといえる。東南アジアの食事傾向においてスープが上位に入っており、図 7 のような汁に野菜や肉を煮た料理が多く見られた。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。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。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 8 のような茶色のコーヒ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。

5.2.2 食事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、その料理における各地域の割合を表 4 にまとめた。その中で、全画像における地域内の画像割合よりも 10% 以上高いいくつかの料理カテゴリに関して注目した。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。これは、東アジアにおいて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。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であると考えられる。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。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ける地域分布に影響を与えているのが分かる。

表 4 食事別における地域傾向

料理カテゴリ	東ア	東南ア	北ア	南ア	ヨー	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	57.40	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	60.40	35.93	0.00	3.66	0.00	0.00
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

東ア：東アジア、東南ア：東南アジア、北ア：北アメリカ

南ア：南アメリカ、ヨー：ヨーロッパ、中東：中東

5.3 特徴量の可視化

最後に食事画像のどのような箇所に地域特徴量が表れているのかを調べた。Grad-CAM[7]には学習を行ったモデルが必要なため、出力層を分析を行った6地域としたVGG16[8]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し用いた。本研究では、肉料理とパン類に絞り実験を行った。図4と図5は実際に地域特徴箇所が表示された例で、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4は、東南アジアにおける肉類の画像で、丸い部分が特徴的であり、肉の丸い部分に地域特徴が表れた。図5は、中東におけるデザートの画像で、バクラバアといわれるデザートでパイ生地の色緑色の部分に地域特徴が表れた。このように地域特有な形状の発見を行うことができたが、今回発見できた地域特徴が表れている画像が少ないためより適切な分析手法が必要となる。

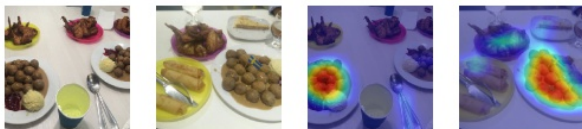


図 4 東南アジアにおける肉類の画像に対する Grad-CAM の適用例

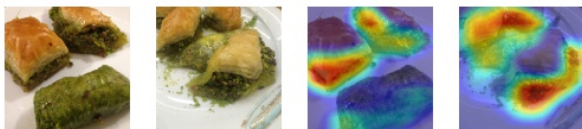


図 5 中東のデザートの画像に対する Grad-CAM の適用例

6. まとめと今後の課題

本研究では、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食事画像における地域傾向の発見を試みた。分析の結果から、地域別に独自の食事画像傾向や、食事画像ごとに割合を占める地域を確認することで、地域における食事傾向を発見することができた。また、地域特徴を表している部分の発見も行った。課題として、Twitter上に投稿されている食事画像が極端に少ないため、南アジア、アフリカ、オセアニアの3地域について分析出来なかった点が挙げられ、Twitter以外の食事画像収集法が必要である。その他、より適切な分析のため、クラスタ内の適切な

ノイズの削除を行うことも課題となる。



図 6 東アジアにおける麺類の画像例



図 7 東南アジアにおけるスープの画像例



図 8 中東における地域特有の飲料の画像例

参考文献

- [1] He, K., Zhang, X. and Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *Proc. of IEEE Computer Vision and Pattern Recognition* (2015).
- [2] Jing, Y. and Baluja, S.: VisualRank: Applying PageRank to Large-Scale Image Search, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 11, pp. 1870–1890 (2008).
- [3] Krizhevsky, A., Sutskever, I. and Hinton, G.: ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems* (2012).
- [4] Lukas, B., Matthieu, G. and Luc, V.: Food-101 – Mining Discriminative Components with Random Forests, *Proc. of European Conference on Computer Vision* (2014).
- [5] Matsuda, Y., Hoashi, H. and Yanai, K.: Recognition of Multiple-Food Images by Detecting Candidate Regions, *Proc. of IEEE International Conference on Multimedia and Expo (ICME)* (2012).
- [6] Nagano, T., Ege, T., Simoda, W. and Yanai, K.: A Large-scale Analysis of Regional Tendency of Twitter Photos Using Only Image Features, *Proc. of IEEE International Conference on Computer Multimedia Information Processing and Retrieval* (2019).
- [7] Selvaraju, R., Cogswell, M. and Das, A., Vedantam, R., Parikh, D. and Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, *Proc. of IEEE Computer Vision and Pattern Recognition* (2016).
- [8] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *CoRR*, Vol. abs/1409.1556 (2015).
- [9] Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J. Chen, B. and Wu, Y.: Learning finegrained image similarity with deep ranking, *Proc. of IEEE Computer Vision and Pattern Recognition* (2014).
- [10] 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).
- [11] 川久保秀敏, 樋爪和也, 柳井啓司: GeoVisualRank を用いた単語概念の地域性の分析, 画像の認識・理解シンポジウム (MIRU) (2011/07).
- [12] 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).
- [13] 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).