

Twitter 画像に対する地域別画像タイプの大規模分析

長野 哲也^{1,a)} 會下 拓実^{1,b)} 柳井 啓司^{1,c)}

1. はじめに

今日, Facebook, Instagram, Twitter といった SNS に写真を投稿することは世界中において人々の生活の一部となっており, 旅行など特別なイベントの写真ばかりでなく, 日々の通常の生活のなかで撮影した大量の写真が SNS には存在している. 従来の SNS を含む Web 上の写真の分析は, その量の多さからテキストで画像を絞り込んだ上で, 画像特徴量を用いて画像の内容に関する分析をすることが一般的であった. そうした分析は, より多くの人に閲覧されるために検索可能になるように, 写真内容に関連したタグを付与することが一般的である Flickr のような写真共有サイトではある程度有効であった. しかしながら, SNS では検索されることは目的とはしておらず, フォロワーや友人と体験したことを共有するための手段として写真を投稿することが一般的であり, そのため通常は一目で内容が分かる写真に対して, あえてそれを説明するようなテキストを付与することはほとんど行われないういってよい. したがって, SNS の画像の傾向を分析するにはテキスト情報で絞り込みをせずに, すべての画像に対して画像情報を用いて分析することが望ましい.

従来は, 100 万枚, 1000 万枚といった大量の画像の分析は大規模な計算設備が必要で, 容易ではなかった. しかしながら, 深層学習および畳み込みネットワーク (Convolutional Neural Network, CNN) の登場により, GPU を利用することで, 大量画像に対する認識や特徴抽出を高速に実行することが可能となった. 実際, 本研究では 1 分間に約 1000 枚の画像からの特徴抽出を行い, 1 台のシングル GPU 計算機で 100 万枚の画像特徴抽出を 17 時間程度で行うことができた.

そこで, 本研究では, CNN による高速特徴抽出を用いて, 大量の Twitter 画像の分析を行うことを目的とする. 特に, 位置情報付きの画像ツイートを対象として, 地域ごとの画像ジャンル分類を行い, 地域による投稿画像の違いを発見することを目的とする.

2. 関連研究

Twitter 画像の分析は, テキスト解析と画像認識を併用し, イベント画像検出 [5][4] などが行われてきた. [4] は, ジオタグと写真の両方を含むツイートを用いて, イベント

を検出する手法を開発した. 大量のデータに対応するため, テキストを分析し, イベントに関するキーワードの検出を行い, 画像をクラスタリングし画像を分類した. 実験結果では, 花火などのイベント画像が検出され, イベント画像の適合率は 65.5 % となった.

しかし, これらの研究では, テキスト解析でツイートを絞り込むため, 画像のみの投稿や, テキスト解析で当てはまらない投稿が捨てられている問題点があった.

そこで, 本研究では, テキストを使わずに画像特徴のみで Twitter 画像を数百万枚単位で分析することで, 地域による投稿画像の傾向の違いの発見を目的とする.

3. 手法の概要

本研究では, Twitter 上の画像を集め, 画像の特徴量を取り出し, 取り出した特徴量でクラスタリングを行い画像を分類する. そして, 分類された画像を時間や場所のデータから分析を行うためのシステムを作成し分析を行う. 手法の概略図を図 1 に表す. ステップ (a) では画像を収集し, 画像の特徴量を取り出す. ステップ (b) では画像の特徴量からクラスタリングを行う. ステップ (c) では分類された画像を画像のジャンルについて区別する. ステップ (d) ではジャンル分けされた画像を時間と場所のデータに基づき地域別の特徴を分析する.



図 1 本研究の流れ

4. 手法の詳細

4.1 画像の収集

我々の研究室では, 2011 年から現在に至るまでの位置情報付き画像の Twitter ログを収集している. 本研究では, 2016 年の 1 月から 6 月までの期間のログを使用し画像を収集する.

4.2 画像の分類

4.2.1 特徴抽出

収集した画像を, 1000 種類, 画像 100 万枚で事前学習された畳み込みネットワークである DCNN 特徴 [2] を利用し, 画像の特徴量抽出を行う. この特徴量の使用によ

¹ 電気通信大学情報理工学部総合情報学

a) nagano-t@mm.inf.uec.ac.jp

b) ege-t@mm.inf.uec.ac.jp

c) yanai@cs.uec.ac.jp

り、画像の意味を考慮したクラスタリングを行うことが可能となる。DCNN 特徴量の抽出には Caffe[3] を使用した。Caffe[3] を使用し、GPU を用いて画像から特徴抽出を行った。これによって、1 分あたり約 1000 枚の高速な特徴抽出が可能となった。大量の画像の分析においては高速に特徴を抽出できることが重要である。Caffe の事前学習済みのモデルとして、本研究では AlexNet をベースとした CaffeNet を使用した。

4.2.2 クラスタリング

本研究では、クラスタリングの手法において一般的な K-means を採用した。抽出した画像の特徴量で画像を分類した。DCNN 特徴量を用いてクラスタリングを行うことで、大量の画像を意味的なまとまりに分類することが可能となり、画像の内容分析が容易となる。K-means は初めにいくつに分類するか決定する必要がある。本研究では、クラスタ数を複数用意し、それぞれのクラスタ数による結果を出した。結果については図 3 のように表示した。本研究では、Python のオープンソースの機械学習ライブラリの scikit-learn にある K-means を使用した。

4.2.3 次元圧縮

Caffe で抽出した画像の特徴量は 4096 次元となっている。そのまま、分類を行うと計算量が膨大となり時間がかかるので、分類の時間を抑えるため、scikit-learn のライブラリにある PCA を使用し、画像の特徴量の次元の圧縮を行った。抽出した特徴量を PCA を用いて圧縮することが分類に影響を及ぼすかどうか検証した。画像 1000 枚を用意し、特徴量の抽出は Caffe で行い、K-means で分類を行う際に PCA を使用し、画像の特徴量を 128 次元に圧縮し K-means を行った。クラスタリングの数は 50 とした。図 2 が特徴量を圧縮せずに K-means を行い、図 3 が特徴量の圧縮を行い K-means を行った。図 3 の結果として、文章の画像や家の画像、夜景の画像でまとまっていた。

図 2 と図 3 の結果から、圧縮された次元を使用しても、

クラスタ33



クラスタ34



クラスタ35



クラスタ36



図 2 画像の特徴量抽出した結果

クラスタ16



クラスタ17



クラスタ18



クラスタ19



図 3 特徴量を pca で圧縮した結果

遜色なく画像がまとまって分類出来た。よって、画像の特徴量を 128 次元に圧縮して画像の分類を行っても問題は無いと判断した。この結果は、[1] の結果とも整合性が取れている。そして、分類された画像を時間と場所のデータを用いて分析を行う。

4.3 結果の可視化

4.3.1 クラスタリングされた画像の表示

K-means によってクラスタリングされた画像の表示は、図 2、図 3 のように画像を表示する。クラスタリングの結果を保存したファイルを選択し、表示するクラスタ数の範囲を選択できるようにした。

4.3.2 GUI への表示

クラスタリングの結果を受けて、画像がどの地域で投稿されたかを分析するため、Google Maps API を使用した。Twitter のログに投稿された画像の緯度経度が記録されているため、その緯度経度から Google Maps にマーカーを作成した。図 4 が Google Maps 上に表示した例である。

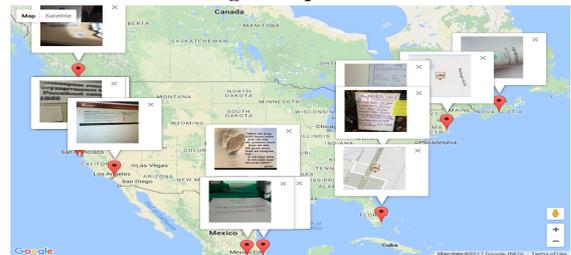


図 4 Google Maps の表示例

5. 実験

5.1 実験データ

実験では、2016 年 1 月から 6 月までの位置情報付き画像 Twitter ログを用い、そのログから位置情報付きで日付のデータを持っているツイートから画像を取り出した。取り出した画像数は合計 216 万 1000 枚となった。今回の実験では、ログから緯度経度、日付、画像名を使用した。緯度経度から投稿された場所を特定した。また、記録されている日付は、日本時間のため時差を考慮し、現地の時間帯に合わせた。ただし、サマータイムは考慮していない。

5.2 クラスタリング

各画像の DCNN 特徴量を Caffe を用いて抽出した。取り出した特徴量を K-means で分類を行う。この際、画像を分類する時の特徴量の重心を取り出した。そして、取り出した重心と各画像の特徴量を計算し、分類を行った。scikit-learn にある K-means と PCA を使用すると、重心と各画像の特徴量の距離の計算結果が出てこない。重心と特徴量の距離で画像をソートするので、重心と特徴量の計算を行った。重心を取り出す際の画像は、各月から画像をランダムで 1 万 7000 枚を取り出し、計 10 万 2000 枚で重心を抽出した。抽出した重心の数（クラスタリング数）は 100, 200, 500 個取り出した。この 3 つの重心を使用し、クラスタリングを行った結果から、地域間の共通の画像のジャンルを分析する。分析結果から、クラスタリング数を増やしても共通の画像のジャンルが変化しなかったのでクラスタリング数を 100 にして各地域の分析を行った。

5.3 分析方法

分析した地域は、図 5 の東アジア、北米、南米、ヨーロッパ、アフリカ、中東、南アジア、東南アジア、オセアニアの 9 つの地域に定め、各地域の画像の特徴について分析を行った。(中国では、Twitter が規制されているので、東アジアから除外。) 分類された画像について、画像のジャンルの分析を行った。分析を行う際に、分類された各クラスごとの画像の数を集計した。集計の際に、重心と各クラスに含まれる画像の特徴量の距離の平均を取り平均以上のものはカウントせずに集計を行った。これを行うことで、重心から遠い画像を排除でき、種類分析の質が向上する。そして、画像の数が多きクラスに注目し、そのクラスに含まれる画像のジャンルを分析した。また、画像の表示の際には、画像の特徴量が重心に近い順でソートを行い画像を表示した。画像を集計した数を指標にし、重心の数や各月、地域別で画像のジャンルについて分析を行った結果、9 つの地域で検出された共通の画像のジャンルとしては、人、建物、文章、風景、食べ物の 5 つのジャンルであった。

共通の画像の 5 種類を使用し、各地域の特徴を分析した。各地域の分析を行うにあたって、画像の投稿された数が多い画像のジャンルに焦点を当て分析を行った。そうすることで、その地域で多く投稿される画像のジャンルと数が判明し、そこから、その地域の特徴を見いだせると考えた。そして、本実験では、画像の投稿数が少ない(100 枚以下) クラスは画像のジャンルに関係なく投稿数の切り捨てをすることにした。理由としては、投稿数が少ないということは、その地域における特色を表していないと判断したためである。そして、5 ジャンルの画像のみの投稿数の集計を行った。各クラスを 5 ジャンルのいずれかに手動分類することによって 5 ジャンルの画像のみの投稿数の集計集計を行った。5 ジャンル分類は自動でも可能であるが、今回は分析精度を重視して、手動で行った。クラスによっては、5 ジャンルのどれにも該当しない場合、複数ジャンルが混ざる場合、ほぼおなじ画像が大量に占める場合、など分析には不適切なクラスがあり、自動では除去が難しい場合があり、最終的に手動で分類した。また、地域別に画像が投稿された時間にも着目して分析を行った。時間の分類は、朝(午前 4 時～午前 10 時)、昼(午前 10 時～午後 15 時)、夕方(午後 15 時～午後 18 時)、夜(午後 18 時～午前 4 時) で分類を行った。分析した結果として、1 月から 6 月までの分析結果の合計を表 1 に記した。そして、1 月から 6 月までの地域別の時間帯の投稿割合を表 2 に記した。また、1 月から 6 月までの各地域の画像の割合を図 6 で表した。

5.4 地域別分析

分析結果の図 6 や表 1 から、東アジアは人の画像がほとんど無く、建物や食べ物(図 7) を多く投稿していた。建物と食べ物の割合を合わせると投稿数のうち 7 割を占める結果になった。北米は、人、建物(図 8) の画像の割合が多かった。南米は、人の画像を多く投稿しており、他の画像の種類の投稿と比べて倍以上の差をつけて投稿されていた(図 9)。ヨーロッパは、9 つの地域の中で投稿数がトップであった。全種類がバランスよく投稿されていた。アフリカ

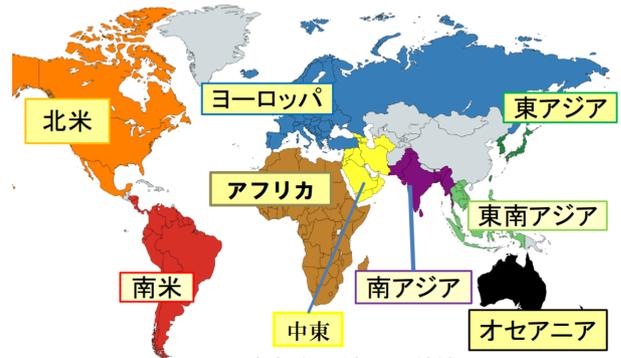


図 5 本実験で分析した地域

は、建物、風景、食べ物の画像がほとんど投稿されており、人の画像が投稿数の 7 割を占めていた。中東は、ヨーロッパより投稿数が少ないものの、全種類がバランスよく投稿されていた(図 10)。南アジアは、全体の投稿の割合を文章が半数を超えていた。これは、他の地域には見られなかった。東南アジアは、南米やアフリカと同様に、人の画像が多かったが、それに加えて、食べ物の画像も多かった。全体の各地域の食べ物の画像数でみると、東南アジアが一番多く食べ物の画像を投稿していた。オセアニアは、全地域で一番投稿数が少なかった。人や文章、風景の画像の割合が多かった。そして、オセアニアでは地図が多く投稿されていた。今回は、地図も文章としてカウントしたため、オセアニアでは文章の画像の割合が多い結果となった。

5.5 時間別分析

時間別の分析において、表 2 を見ると東アジアの投稿の時間帯の特徴として、朝に投稿されるパーセンテージが約 1 割で、夜に投稿する割合が約 5 割であった。朝の投稿の割合が 1 割だったのは、この地域以外には見られなかった。北米や南米は夕方を除いて、朝昼夜まんべんなく投稿していた。ヨーロッパやアフリカ、中東、南アジア、オセアニアは朝と夜に多く投稿されていた。東南アジアは朝約 2 割、夜約 5 割といったような朝夜の時間帯で投稿されていた。

画像のジャンルによって、投稿した時間帯の差はあまり見られなかった。本実験で検出されたのは、東アジアで、建物と風景の画像が昼と夕方投稿が多く、東南アジアで、食べ物の投稿が夜に多かったのみであった。

表 1 1 月から 6 月までの各地域の分析結果の合計

国	投稿数	人	建物	文章	風景	食べ物
東アジア	65,530	459	20,172	8,973	8,217	27,709
北米	124,141	40,231	38,336	12,015	11,314	22,245
南米	76,481	51,471	3,613	16,416	2,381	2,600
ヨーロッパ	183,250	63,942	37,651	34,511	22,018	25,128
アフリカ	24,101	17,409	284	6,408	0	0
中東	44,184	11,681	10,216	12,334	4,123	5,830
南アジア	13,263	5,424	277	7,214	295	53
東南アジア	145,088	73,265	15,862	13,120	4,323	38,518
オセアニア	4,695	1,049	144	2,214	953	335

6. 考察

今回の実験では、計 216 万 1000 枚の画像から、設定した地域における共通の画像ジャンルを検出出来た。そして、得られた共通の 5 ジャンルの画像から地域の特徴を発見出

表 2 1 月から 6 月までの地域別の時間帯の投稿割合

国	朝	昼	夕方	夜
東アジア	11.48%	19.93%	18.62%	49.97%
北米	28.95%	23.57%	13.70%	33.78%
南米	28.33%	24.30%	12.67%	34.72%
ヨーロッパ	32.13%	12.88%	11.83%	43.13%
アフリカ	34.43%	15.20%	12.27%	38.13%
中東	36.80%	18.63%	7.18%	37.42%
南アジア	31.17%	16.73%	9.27%	42.83%
東南アジア	22.70%	12.97%	11.12%	53.18%
オセアニア	26.40%	15.98%	13.62%	43.98%

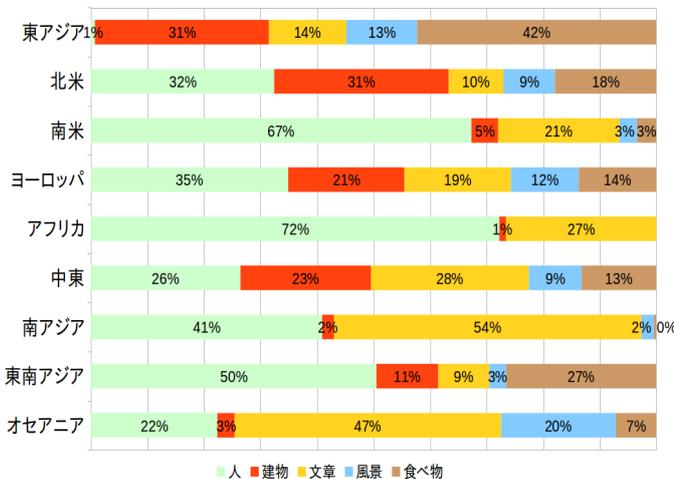


図 6 1 月から 6 月までの 5 種類の画像の地域別割合

来た。しかし、今回の実験では、画像の投稿数が少ないクラスを切り捨てて、分析結果に加えなかったため、アフリカには風景や食べ物の画像が全くないことが生じたと考えられる。また、オセアニアの文章の画像は、地図も文章とカウントしたため、オセアニアは全体的に文章の割合が多くなったと思われる。

全体の地域別の割合を見ると、東アジアと東南アジアでは、食べ物の画像の割合が多く、南米やアフリカ、東南アジアでは、人の画像が 5 種類の合計の投稿数の半数を超えるといったような地域間における共通の特徴が検出された。画像の投稿数が違うが、ヨーロッパと中東は全種類がバランスよく投稿されている共通点を発見した。また、東アジアは人の画像の投稿がほとんど無いことや、南アジアでは文章の投稿が 50% で超えているという地域独自の特色が明らかになった。画像が投稿された時間帯を調べると、東アジアの建物と風景の投稿の時間帯が昼と夕方が多かったこと以外では、各地域で画像の種類に関係なく、ある一定の時間帯で画像を投稿していることが分かった。こういったことから、東アジアと東南アジアでは、食べ物に関心を持っていることや南米やアフリカ、東南アジアでは、人の画像を投稿することを一種の楽しみとしているのではないかと考えられる。投稿された時間帯から、それぞれの地域における、人の行動パターンを区別出来ると考えられる。

7. まとめ

本研究では、画像に関するメタデータを持つ Twitter の画像を収集し、GPU を用いて高速に大量画像から DCNN 特徴量の抽出を行ったあと、K-means によるクラスタリングを行った。そして、クラスタリングの結果から地域間

の共通の画像ジャンルを分析した。共通の画像のジャンルと画像に関する日付と緯度経度を使用し、地域間の画像の特徴の違いや投稿の時間帯の分析を行った。分析の結果から、地域別の画像の特色や投稿の時間帯が地域によって傾向があることが見られた。

8. 今後の課題

地域別の特徴を見出すために、共通の画像ジャンルを特定したことで、その地域にしかない画像ジャンルを分析することが出来なかったと思われる。画像の種類を特定せずに分析を行うことで、地域独自の画像を発見することが出来るようになる。また、本実験で定めた 5 種類の共通画像のジャンルを細分化することで、新たな特徴を発見出来ると思われる。例えば、人の画像を細分化し、自撮りか集合写真かどうかのようにジャンルを細かくすることで、地域の新たな特徴が判明されると考えられる。



図 7 食べ物の例 (東アジア)



図 8 建物の例 (北米)

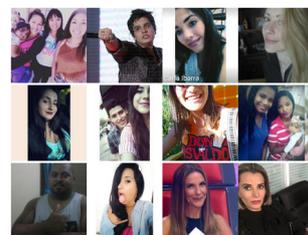


図 9 人の例 (南米)



図 10 文章の例 (中東)

参考文献

- [1] Babenko, A., Slesarev, A., Chigorin, A. and Lempitsky, A. V.: Neural Codes for Image Retrieval, *Proc. of European Conference on Computer Vision*, pp. 584-599 (2014).
- [2] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E. and Darrell, T.: DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition., *Proc. of International Conference on Machine Learning*, pp. 647-655 (2014).
- [3] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T.: Caffe: Convolutional architecture for fast feature embedding, *Proceedings of the 22nd ACM international conference on Multimedia*, ACM, pp. 675-678 (2014).
- [4] Kaneko, T. and Yanai, K.: Event photo mining from twitter using keyword bursts and image clustering, *Neurocomputing*, Vol. 172, pp. 143-158 (2016).
- [5] Schinas, M., Papadopoulos, S., Mitkas, P. A. and Kompatsiaris, Y.: Visual Event Summarization on Social Media using Topic Modelling and Graph-based Ranking Algorithms, *Proc. of International Conference on Multimedia Retrieval*, pp. 203-210 (2015).