

Twitterにおけるアイコンの色彩的特徴と投稿の感情極性の関連分析

Analysis of Relationship Between Color Characteristics of Icons and Emotional Polarity of Posts on Twitter

社領 一樹^{*1}, 金城 篤史^{*1}, 鈴木 大作^{*1}
 Kazuki SHARYO^{*1}, Atsushi KINJO^{*1}, Taisaku SUZUKI^{*1}
^{*1} 沖縄工業高等専門学校
^{*1}National Institute of Technology, Okinawa College
 Email: ac214702@edu.okinawa-ct.ac.jp

あらまし：近年、その拡散力の高さから個人・団体による SNS を利用した商品・活動の宣伝が盛んになっている。一方で投稿内容によっては「ネット炎上」と言われる状況に陥るリスクもある。これを抑制する対策の1つとして、ネガティブな投稿を行わないことを意識することがあげられる。このことから、色の持つ効果と実体験から「アイコンの色によって投稿内容が変化するのはないか」という仮説を立てた。本研究では日本において最もメジャーな SNS である Twitter を例にとり、アイコンの色と投稿(リプライ)の感情極性の相関関係についての分析を行い、相関性がみられるという結果を得た。

キーワード：アイコン画像, 色, 感情極性, Twitter

1. はじめに

SNS アカウントの運営には、フォロワーの獲得やネット炎上回避のために、ネガティブな投稿を避ける必要がある。また、色には見た人の感情を変化させることや、統計的に好かれやすい色、好かれにくい色があることが先行研究^(1,2)によって示されている。これらのことから、「アイコンの色によって投稿およびそれに対するリプライの感情極性が変化するのはないか」という仮説を立て、実際に相関があるのであれば、ネガティブな投稿を抑制する効果のある色を示すことで SNS の運営戦略に役立てられると考えた。

そこで本研究では、日本で最もユーザ数が多い SNS と言える Twitter を例に、ユーザのアイコンの色と投稿の感情極性を分析することで、それらの相関関係の有無を明らかにすることを目的とする。類似の研究として、Twitter のアイコンを 13 種のカテゴリに分類し各カテゴリに該当するユーザデータを 100 人ずつ収集し、ユーザ行動について分析した研究があるが⁽³⁾、投稿内容について分析していないことやカテゴリに該当するユーザの特性に偏りが生じていることなどの課題が挙げられる。本研究では、クラスタリングによってアイコンからメインカラーと取り出し、それをあらかじめ用意した 36 色に分類し、それと全体に占めるポジティブな投稿(リプライ)の割合を決定木、連関規則によって分析する。

2. 研究手法

2.1 データ収集

Twitter API を用いてユーザのアイコン・ツイート・他のユーザからのリプライ・フォロワー数・フォロワー数を取得する。サンプルユーザとして 1142 人分のデータを収集した。

2.2 アイコン色の分類

収集したアイコンのメインカラーを取り出すため、クラスタリングを行う(図 1)。本研究では、Python の scikit-learn を用いた K-means 法によるクラスタリングを行った。次に、取り出した色についてあらかじめ用意した色に分類を行う。今回は、JIS 規格 S6028:2007 に基づく水彩絵の具 36 色とした。分類は次のような手順で行う。

1. $dif=765(=255 \times 3)$ とする。
2. 各 RGB 値の差, dif_r, dif_g, dif_b を算出する。
3. 各 RGB 値の差の合計を dif_sum とする。
($dif_sum=dif_r+dif_g+dif_b$)
4. dif_sum の平均値を $dif_average$ とする。
($dif_average=dif_sum/3$)
5. 「各 RGB 値の差」と「差の平均値」の差の合計値を dif_tmp とする。
6. 36 色と比較する過程で dif_tmp が dif より小さい場合には, $dif=dif_tmp$ とする。
7. すべての処理を終えた段階で dif が持っている値の色に分類する。

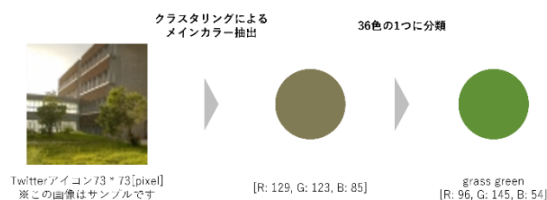


図 1 アイコン色のクラスタリング・分類手順

2.3 投稿(リプライ)の感情極性分析

Python ライブラリである asari を用いて、ユーザの投稿および他ユーザからのリプライの感情極性を分析する。投稿の positive または negative の判定に従いそれぞれカウントを行い、すべての投稿に占める positive 投稿の割合 $tweet_positive_ratio$ (リプライの場合には同様の手順で $reply_positive_ratio$) を求める。

2.4 入力データの作成

収集したデータをもとに, screen_name (Twitterにおける一意な名前), color (分類した36色), tweet_positive_retio, reply_positive_retio を列に持つCSVデータを作成する。

3. 分析と結果

R Studio のライブラリ rpart を用いて決定木による分析を行う。color を従属変数とし, tweet_positive_retio, reply_positive_retio を説明変数としたときの分類木の結果を示す。complexity parameter の値 (以降 cp 値と表記) は, 0.002 に設定した。その結果を図2に示す。

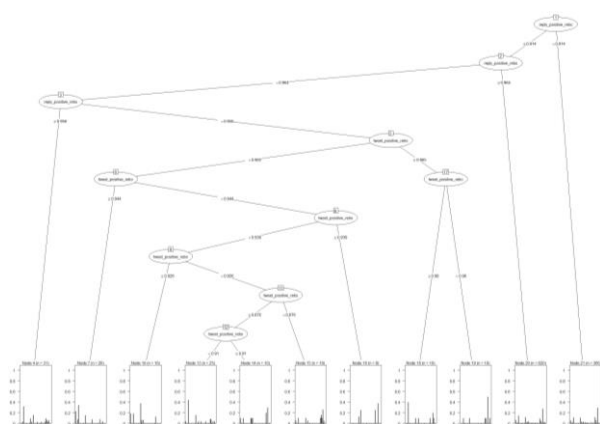


図2 color を従属変数, cp = 0.002 としたときの分類木

reply_positive_retio が根に近いブランチとして現れたことがわかる。特に, $0.956 \leq \text{reply_positive_retio} < 0.964$ の範囲では burnt_umber, それ以外の範囲では white の突出が目立った。特に, $0.956 \leq \text{reply_positive_retio} < 0.964$ の範囲では burnt_umber, それ以外の範囲では white の突出が目立った。

次に, tweet_positive_retio を従属変数としたときの結果を図3示す。cp 値は 0 とした。

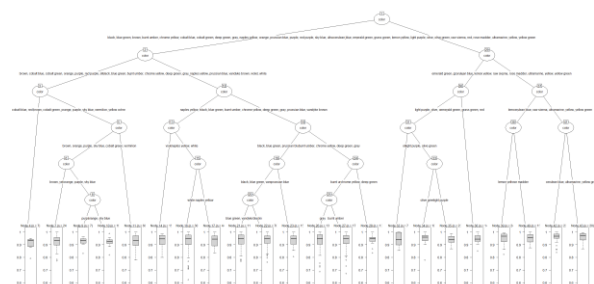


図3 tweet_positive_retio を従属変数, cp = 0 としたときの分類木

このとき, 根の分岐で2つに分けられた色を, 36色の辞書データに着色を施して示した(なお, green, viridian は出現しなかった)(図4)。分類された色に

まとまりがあることが観察できる。白背景で示した辞書データの色は, そうでない色に比べて tweet_positive_retio の最大値および中央値が概ね高いことがわかった。

Rose madder	[241, 91, 134]	Emerald green	[0, 168, 105]
Red	[206, 34, 60]	Cobalt green	[0, 164, 129]
Vermilion	[222, 56, 56]	Deep green	[27, 81, 63]
Brown	[148, 73, 61]	Viridian	[0, 101, 73]
Vandyke brown	[91, 68, 49]	Blue green	[0, 123, 122]
Raw sienna	[177, 99, 42]	Sky blue	[46, 152, 204]
Orange	[255, 113, 68]	Cerulean blue	[0, 130, 181]
Burnt umber	[94, 66, 52]	Cobalt blue	[0, 96, 175]
Yellow ochre	[192, 132, 42]	Prussian blue	[42, 61, 97]
Chrome yellow	[250, 190, 0]	Ultramarine light	[82, 112, 201]
Naples yellow	[254, 205, 121]	Ultramarine	[14, 63, 158]
Yellow	[244, 213, 0]	Violet	[71, 59, 135]
Olive	[129, 121, 35]	Light purple	[159, 90, 174]
Lemon yellow	[238, 234, 0]	Purple	[97, 52, 116]
Olive green	[85, 101, 45]	Red purple	[194, 42, 119]
Yellow green	[101, 192, 66]	White	[234, 234, 234]
Grass green	[96, 145, 54]	Gray	[156, 156, 156]
Green	[0, 140, 75]	Black	[52, 52, 52]

図4 36色の辞書データに分類に基づく着色を施した図

4. 考察

決定木の結果から, tweet_positive_retio, reply_positive_retio と color との間に相関関係がみられることがわかった。特に, tweet_positive_retio を従属変数, color を説明変数とした際の結果から, 色彩傾向が投稿の感情極性に影響することが示され, yellow から emerald green の範囲では投稿内容が positive になる傾向があると考えられる。

一方で, 今回の分析に用いたデータでは color の出現回数が不揃いであったことや, 1142人のデータを36色に振り分けた結果, 1つのカテゴリにつき平均31人程度しかデータを集められていないことなど, 入力データの妥当性の検証が必要であると考えられる。

5. まとめ

本論文では, SNSのアイコンの色と投稿(リプライ)内容の関係に着目し, Twitter APIを用いたデータ収集を行い, 決定木による相関性の有無についての分析を行った。その結果, アイコンの色と投稿(リプライ)の感情極性に相関性があることが示された。しかし, 入力データが妥当なものであったかの検証が必要であると考え, 今後は color の色数を削減しそれぞれの色の出現数を揃えたデータでの分析を行っていく。

参考文献

- (1) 仁科恭徳: "若者の色彩感覚に関する実態調査", カルチュール, 9(1), pp55-62 (2015)
- (2) 稲浪正充・栗山智子・安部美恵子: "色と感情について", 島根大学教育学部紀要(人文・社会科学), 第28巻, pp35-50, (1994)
- (3) 富永登夢・土方嘉徳: "Twitter上のアイコン画像とユーザ行動の関係の調査と分析", インタラクシオン2016論文集, pp272-277, (2016)