

Twitter におけるアイコン画像の色彩とツイートの感情極性値の相関

Correlation Between Icon Image Color and Emotion Polarity of Tweets 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: SNS は拡散力に優れ, 多くの個人・組織・団体が活用する一方で, ネガティブな投稿によりフォロワーの減少やネット炎上に陥るリスクがある. SNS では, アカウントにアイコン画像を設定することができ, ユーザが個性を表現するための1つの手段である. 色彩は人間の心理に影響を与えることが知られているが, アイコン画像の色彩についても投稿者の感情に影響を及ぼすと考えられる. 本研究では, Twitter のユーザを対象に, アイコン画像の色彩とツイートの感情極性値の関係性について分析を行い, 相関があるという結果を得た.

キーワード: Twitter, アイコン画像, 色彩, 感情分析, 相関分析

1. はじめに

近年, 情報拡散力に優れる SNS を商品や活動の発信に活用する個人・組織・団体が増えている. 一方で, SNS にはネガティブな投稿によりフォロワーの減少やネット炎上という状況に陥るリスクもある. また, 色彩には好まれやすさやポジティブ・ネガティブな言葉のイメージがあることが, 仁科⁽¹⁾や稲浪⁽²⁾の研究で示されている. 多くの SNS では, アカウントのプロフィールにアイコン画像を設定できることから, アイコン画像の色彩が投稿者の感情に影響を及ぼし, 投稿内容に変化を生じさせることが考えられる. アイコン画像の色彩と投稿の感情極性に有意な関係があれば, ネガティブな投稿を抑制する効果がある色を示せる. これは, SNS アカウントの運用の一助になると考える.

本研究では, Twitter の日本語投稿を行うユーザを対象に, アイコン画像の色彩とツイートの感情極性値の関係性について分析する.

2. 手法

2.1 データの収集・前処理

Twitter REST API を用いて, 日本語投稿を行う Twitter ユーザのアイコン画像, ツイートを収集する. 次に, 取得したツイートについて, ハッシュタグや URL などの特殊文字列を除外, 2文字未満の投稿の除外, 日本語が含まれない投稿の除外を行う. さらに, 1つのアイコン画像が使われる期間に行われるツイートを180件と考え, 各ユーザの最新ツイート最大180件のみを選定する. この処理を経て, 有効なツイートが残っていないユーザと日本語以外のツイートのみユーザを除外し, 残ったユーザを分析対象とした. これにより, 有効データは74,461人のユーザ, 11,020,490件のツイートとなった.

2.2 アイコン画像の主要3色の取り出し

取得した各ユーザのアイコン画像について, K-means++法による画素値のクラスタリングを行い, 主要な3色の取り出しを行う. 次に, 取り出した3色の画素値について, 人間が理解しやすい色相・彩度・明度のパラメータであらわすために, RGB カラーモデルから HSV 色空間へ変換する.

2.3 ツイートの感情極性値の算出

取得した各ユーザのツイートについて, asari, BERT, oseti という3つの感情分析ツールを用いて, ポジティブ度を0~1の値であらわす感情極性値 (positive ratio, 以降 pr 値)の算出を行う.

asari は, Python の日本語の感情分析パッケージで, 文章の感情極性値をポジティブラベル, ネガティブラベルの両方について0~1のスコアで返す. ここでは, スコアを式(1)に基づいて pr 値とする.

$$pr = score_{asari}(label: positive) \quad (1)$$

BERT は, Transformer ベースの自然言語処理の事前学習用モデルである. ここでは, 日本語感情分析用モデル「daigo/bert-base-japanese-sentiment」(以降 daigo/bert)を用いる. daigo/bert は, 文章がポジティブかネガティブか判定し, 0.5~1のスコアを返すため, スコアを式(2)に基づいて pr 値とする.

$$pr = \begin{cases} score_{bert}(label: positive) \\ 1 - score_{bert}(label: negative) \end{cases} \quad (2)$$

oseti は, 日本語極性辞書を用いた Python の日本語感情分析ライブラリである. 辞書には, 東北大学の乾・鈴木研究室が公開している「日本語評価極性辞書(用言編)」⁽³⁾および「日本語評価極性辞書(名詞編)」⁽⁴⁾を用いる. oseti は, 文章の感情極性を句点区切りの1文ごとに-1~1のスコアで返すため, n個の句点区切りの文章についてスコアを式(3)に基づいて pr 値とする.

$$pr = \frac{1}{n} \sum_{i=1}^n \{(1 + score_{oseti_i}) \div 2\} \quad (3)$$

asari, daigo/bert, oseti で算出したツイートごとの pr 値の平均値をとり、ユーザごとの pr 値を決定する。

2.4 相関分析

取得したアイコン画像の主要な3色のHSVパラメータ(h1, s1, v1, h2, s2, v2, h3, s3, v3)と、ユーザごとの pr 値について、決定木分析と連関規則分析を行う。

3. 分析結果

3.1 決定木分析

R 言語の rpart ライブラリを用いて、CART Regression による回帰木を作成した(図 1, 図 2, 図 3)。目的変数は pr 値で、説明変数はアイコン画像の主要な3色のHSVパラメータである。いずれの場合においても、根の分岐に明度のパラメータが出現し、閾値を基準に明度が明るい側に分類されたアイコン画像のユーザの pr 値が高い傾向があった。

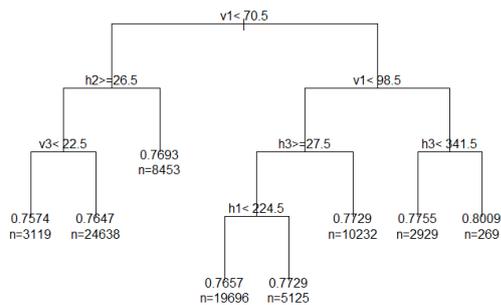


図 1 回帰木の結果

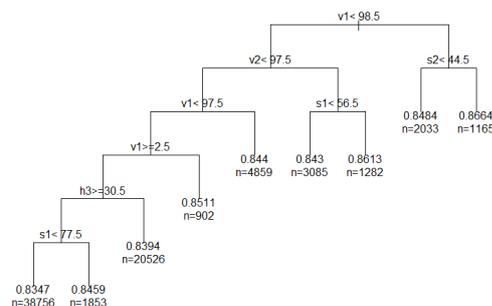


図 2 回帰木の結果(daigo/bert)

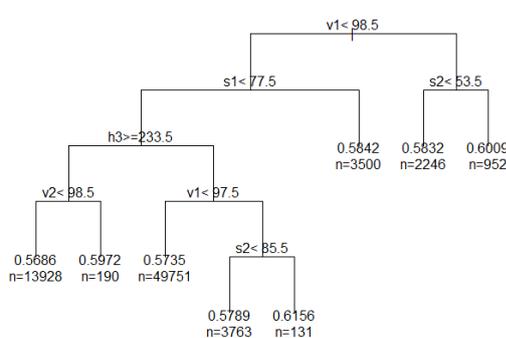


図 3 回帰木の結果(oseti)

3.2 連関規則分析

R 言語の arules ライブラリを用いて、Apriori による連関規則分析を行った。ルール条件部には、アイコン画像の主要な3色のHSVパラメータ、結論部には pr 値を指定した。結果を表 1 から表 3 に示す。無彩色、彩度が低く明度が高い色、彩度と明度が高い色のときや、彩度が低く明度が高い色と彩度が高い色の組み合わせのときに pr 値が高い傾向があることが読み取れる。

表 1 連関規則の結果(asari)

条件部	結論部	支持度	信頼度	リフト値
v1=[0,54] s2=[10,31]	[0.00586,0.731]	0.04089389	0.3519011	1.055717
v1=[86,100] v3=[75,100]	[0.809,1]	0.04472140	0.3516739	1.054993
h1=undefined s1=[0,8]	[0.809,1]	0.03686494	0.3510679	1.053175
v1=[0,54] v3=[0,52]	[0.00586,0.731]	0.03883912	0.3500363	1.050123

表 2 連関規則の結果(daigo/bert)

条件部	結論部	支持度	信頼度	リフト値
s1=[0,8] v1=[86,100] s2=[31,100]	[0.882,0.994]	0.03268825	0.4052614	1.215752
s1=[0,8] s2=[31,100]	[0.882,0.994]	0.03797961	0.3866557	1.159936
h1=undefined s1=[0,8]	[0.882,0.994]	0.03933603	0.3746003	1.123771
s3=[33,100] v3=[75,100]	[0.882,0.994]	0.03638146	0.3736552	1.120935

表 3 連関規則の結果(oseti)

条件部	結論部	支持度	信頼度	リフト値
s1=[0,8] v1=[86,100] s2=[31,100]	[0.592,1]	0.03063349	0.3797869	1.139330
s1=[0,8] s2=[31,100]	[0.592,1]	0.03656948	0.3722997	1.116869
v1=[86,100] v3=[75,100]	[0.592,1]	0.04590322	0.3609674	1.082873
s1=[0,8] s3=[13,33]	[0.592,1]	0.03442070	0.3604275	1.081253

4. 考察・おわりに

決定木分析と連関規則分析の結果から、アイコン画像の明度と、主要な色同士の組み合わせが、感情極性値と結びつきがあると考えられる。

本研究では、日本語投稿を行う Twitter ユーザを対象に、アイコン画像の色彩とツイートの感情極性値の関係性について分析を行った。その結果、相関があることが示された。

参考文献

- (1) 仁科恭徳: “若者の色彩感覚に関する実態調査”, 明治学院大学教養教育センター紀要: カルチャー, 明治学院大学教養教育センター, Vol.9, No.1, pp.55-62 (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), 島根大学, Vol.28, pp.35-50 (1994)
- (3) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.3, pp.203-222 (2005)
- (4) 東山昌彦, 乾健太郎, 松本裕治: “述語の選択選好性に着目した名詞評価極性の獲得”, 言語処理学会第 14 回年次大会論文集, 言語処理学会, pp.584-587 (2008)