

Twitterにおける肌色修飾子利用状況の調査

若山 公威†

名古屋外国語大学†

1. はじめに

Unicode 8.0 から肌色修飾子 (skin tone modifier) が加わり, Twitter などの SNS でも利用が広まっている. コードポイント U+1F3FB (Fitzpatrick タイプ 1-2) から U+1F3FF (同タイプ 6) までの 5 種類の肌色修飾子が用意されており, ベースとなる人物に関する絵文字の後に用いることで, 肌の色を変更させることができる. Ljubešić ら[1]は, 世界の地域別絵文字利用状況を調査した結果, 北米, 西ヨーロッパ, ロシア, オーストラリアなどの地域での特徴的な絵文字として, 肌色修飾子タイプ 1-2, タイプ 3 を挙げている. McGill[2]は, 米国における Twitter での肌色修飾子の利用状況を調べ, タイプ 1-2 の割合が少ないことを示している. これら以外に肌色修飾子の利用状況は明らかになっていない.

本研究では, Twitter での肌色修飾子の利用状況を調査した.

2. 使用データ

Twitter Streaming API を用いて, プロフィールの description または name に肌色修飾子を 1 種類のみ含み, ツイート数が 10 以上のユーザをランダムに取得した. 期間は 2017 年 7 月 12 日から 7 月 19 日までである. 本研究では, このプロフィールにおける肌色修飾子タイプにより, ユーザを 5 つのグループに分ける. タイプ 1-2 のユーザは 37.0%を占めたが, タイプ 6 のユーザは 2.5%のみであり, グループごとのばらつきが見られた. 各グループからランダムに 5,000 ユーザを選び, ツイートを最大 3,200 取得した. ただし, 36 ユーザについては, 作業の間のアカウント削除などによりツイートを取得できなかった. ツイート総数は 83,793,030 である.

3. 調査結果

3.1. ツイートでの利用状況

まず, ユーザごとに, リツイートを除いたツイート本文内で肌色修飾子を利用しているかどうか調べた. 肌色修飾子を全く含んでいないユ

ーザは, 全体の 7%であった. 肌色修飾子を利用しているユーザについて, 最も利用している肌色修飾子タイプを調べた. ユーザタイプごとに分類した結果を表 1 に示す. カイ 2 乗検定において有意水準 1%で有意差が認められた. 自身のプロフィールで用いている肌色修飾子をツイート内でも利用する傾向が高いことが分かる.

表 1: ツイート内修飾子タイプ (%)

		ツイート内絵文字修飾子のタイプ				
		1-2	3	4	5	6
ユーザタイプ	1-2	84.5	12.1	2.4	0.5	0.1
	3	24.9	63.5	10.2	1.1	0.3
	4	5.9	10.1	73.2	10.3	0.4
	5	1.5	1.3	13.2	82.9	1.1
	6	10.3	4.5	6.8	21.7	56.7

ツイート内の肌色修飾子がどのような絵文字に付けられているか調べたところ, すべてのユーザタイプにおいて, “folded hands” (U+1F64F) が 1 位になった. これを含めて上位は, 手あるいは腕に関する絵文字が占めている. 合意や感謝, お願いなどの意図を伝える場合にこれらの絵文字が使われるためと思われる.

3.2. プロフィールでの利用状況

プロフィールで肌色修飾子がどのような絵文字に付けられているか調べた. この結果, ツイートでの利用種類と同様に “folded hands” (U+1F64F) が 1 位となった. ツイート内に含まれる絵文字との大きな違いとして, 全タイプ合計の 3 位が “baby angel” (U+1F47C) となっている点があげられる. プロフィールをチェックしたところ, 赤ちゃんがいることを示しているユーザもいるが, 自身のことを示していると思われるものもある.

次に, プロフィール写真と, プロフィールでの肌色修飾子利用状況の関係を調べた. プロフィール写真を Face++ API¹に入力し, 人種 (Asian, Black, White) を求めた. Face++ API により顔が含まれていない, あるいは, 顔が複数含まれている結果となった写真は対象外とした. 全体で, 顔が含まれていない写真は 25.5%,

An investigation of skin tone modifier usage in Twitter

† Kimitake Wakayama

Nagoya University of Foreign Studies

¹ <https://www.faceplusplus.com/>

顔が複数含まれていたのは 9.1%であった。ユーザタイプごとにどの人種に分類されたか表 2 に示す。カイ 2 乗検定において有意水準 1%で有意差が認められた。なお、Face++ API の性能に関して、Chakraborty ら[3]がアノテータ 3 名と比較したところ 79%の一致となっている。

表 2：識別率 (%)

	ユーザタイプ				
	1-2	3	4	5	6
White	56.3	64.2	41.0	23.1	29.6
Asian	36.6	26.4	24.8	19.4	20.1
Black	7.1	9.4	34.2	57.5	50.3

3.3. リプライツイートの返信先ユーザタイプ

リプライツイートについて、その返信先ユーザがどのユーザタイプであるか調べた。具体的には、ツイート内の `in_reply_to_user_id` をもとに返信先ユーザを求め、そのプロフィールに肌色修飾子が 1 種類のみ含まれているユーザを対象とした。全返信先ユーザ数 476,961 のうち、対象ユーザ数は 31,858 であった。ユーザタイプごとに、返信先のユーザ割合をまとめた結果を表 3 に示す。なお、あるユーザから同一ユーザへのリプライは 2 つ以上あっても 1 ユーザとカウントしている。カイ 2 乗検定において有意水準 1%で有意差が認められた。プロフィールの肌色修飾子がタイプ 1-2 から 5 のユーザは、それぞれ同じタイプのユーザ宛のリプライが最も多いことが分かる。タイプ 6 のユーザについては、タイプ 5 宛のリプライが最も多かった。つまり、自分と同じユーザタイプへの返信が多い。返信先としてタイプ 6 の割合が少ないのは、ユーザの絶対数が少ないことと関連していると思われる。

表 3：返信先ユーザのタイプ (%)

		返信先ユーザタイプ				
		1-2	3	4	5	6
送信元ユーザタイプ	1-2	57.7	21.9	12.4	6.9	1.1
	3	24.9	43.2	17.8	12.4	1.7
	4	6.6	8.7	42.1	39.7	2.8
	5	3.3	5.1	31.7	55.6	4.2
	6	5.4	6.4	27.1	44.8	16.4

3.4. ツイートからプロフィール内肌色修飾子の推定

ツイート本文から、投稿したユーザタイプを推定できるか実験を行った。前節までの調査で

対象としたユーザのうち、プロフィールの `lang` が `en` で、ツイートの半数以上が英語であるユーザのみを対象とした。英語かどうか調べるには `langdetect 1.0.7`¹を使用した。ユーザによっては、ツイートに肌色修飾子が 1 つも含まれていない場合もある。実験の対象ユーザ数は 12,783 である。ツイート内の URL とユーザ名を削除した後、1 ユーザから得られるツイート集合を 1 つの文書とみなし TF-IDF を求め、特徴量ベクトルを生成した。このとき、英単語のみの場合、通常の絵文字を加えた場合、さらに肌色修飾子も加えた場合の 3 通りの特徴量ベクトルを作成した。それぞれの特徴量に対して、Support Vector Machine にて 5 分割交差検定による推定を行った結果を表 4 に示す。肌色修飾子を加えることで推定精度が高くなった。3.1 節で調べたとおり、自身のプロフィールで用いている肌色修飾子を、ツイートで用いる傾向が高いためであろう。絵文字については、ユーザタイプごとの利用傾向ははっきりしていないと考えられる。

表 4：推定精度

特徴量ベクトル	適合率	再現率	F 値
英単語	0.41	0.41	0.41
英単語+絵文字	0.42	0.42	0.42
英単語+絵文字+肌色修飾子	0.60	0.60	0.60

4. おわりに

本研究では、Twitter での絵文字の肌色修飾子利用状況を調査した。その結果、プロフィールで利用している肌色修飾子がツイートで利用され、返信も同じタイプのユーザが多いことが分かった。今後はさらにデータを増やして調査を行ってきたい。

参考文献

- [1] Ljubešić, Nikola, and Darja Fišer, A Global Analysis of Emoji Usage. Proc. of the 10th Web as Corpus Workshop, pp.82-89, 2016.
- [2] Andrew McGill, Why White People Don't Use White Emoji, <https://www.theatlantic.com/politics/archive/2016/05/white-people-dont-use-white-emoji/481695/> (2018/01/06 アクセス).
- [3] Abhijnan Chakraborty et al., Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations, Proc. of the 11th International Conference on Web and Social Media, pp.22-31, 2017.

¹ <https://pypi.python.org/pypi/langdetect>