

Twitterの解析による地域特性の可視化手法

A Method for Visualization of Area Characteristics by Twitter Analysis

平沼 康明 坂本 真樹
Yasuaki Hiranuma Maki Sakamoto

電気通信大学大学院 情報理工学研究科

Graduate School of Informatics and Engineering, The University of Electro-Communications

Various studies have focused on area characteristics. Although area characteristics has been analyzed by statistical data, very little attention has been paid to emotions, interests or personality of people in the areas. This study focuses on area characteristics reflecting emotions, interests or personality of people in those areas. In order to grasp such area characteristics, we analyzed microblog text posted at each area. Recently, microblog text with geotag has attracted attention as text reflecting affective aspects of people in the areas. On the other hand, the area characteristics is frequently visualized by colors such as temperature or population. This study uses colors to visualize area characteristics extracted from microblog texts. We deduced color that reflects the affective image of each area by word frequency and word-color vector of words that are contained in microblog text posted at each area. As a result, we have deduced affective area characteristics by color.

1. はじめに

地域特性とは、地域の持つ独自性を指す言葉で、地域と人を分析するための要素として様々な分野で有用とされてきた。従来の研究では、各地域の統計的なデータなどから地域特性を把握する手法が取られていたが、そこで活動する人々の感性(物事を感じる能力, 感受性, 感覚, 感情)やライフスタイルを反映した地域特性の把握は行われてこなかった。

そこで、現在マイクロブログ上の位置情報付きテキストを利用した研究が注目されている。マイクロブログはその即時性から、利用者の興味関心を反映するメディアといえ、位置情報が付加されたマイクロブログの発言は、ユーザが興味関心を持った場所についての発言といえる [1]。そのため、マイクロブログはその地域で活動する人々の興味関心を探り、各地域特性を得る情報源として非常に有用であるとされ、様々な研究が行われている。

一方、地域の情報を可視化する手法として、天気図などといった、地域に色彩を対応付けてマップ化する手法が広く用いられている。これは、色彩がもつ「色彩の感情的作用」によって、視覚的、直感的に地域の状態を把握することが有効であるためである。

そこで本研究では、感性的地域特性の把握・可視化を目的とし、マイクロブログ上の位置情報付きテキストを媒介に、その地域を色彩と対応付けることで、都市の感性的地域特性を色彩として可視化する手法を提案する。テキストから感性イメージを反映した色彩を推定する手法は、[2]の「テキストの感性イメージを反映した色彩推薦システム」の手法を応用する。このシステムは単語と色彩の認知的連想関係に着目しており、テキスト中の出現単語から連想される色彩から、テキスト全体の連想色彩を推定している。本研究の意義として、人々の感性やイメージ、ライフスタイルを反映した地域特性の視覚的、直感的な把握が可能であるという点と、地域特性を色彩で表現することで、屋外広告のデザイン支援が可能である点が挙げられる。

2. 方法

2.1 本手法の概要

[2]の「テキストの感性イメージを反映した色彩推薦システム」の手法を、位置情報付きマイクロブログテキストに応用し、出力された色彩の傾向をみることで、それを各地の地域特性とする。分析対象は横浜、東京、新宿、品川、渋谷、外苑前の6地域とした。この6地域の選定基準は、人通りが多く、2020年の東京オリンピック時に発展が期待されることである。マイクロブログはTwitterを使用した。

2.2 マイクロブログテキストの収集

TwitterAPIを用いて、2014年11月14日、11月28日、12月05日、12月15日時点で6地域の駅から半径500m以内で投稿されたTweetに対して収集を行った。なお、TwitterAPIの仕様で、一度に収集できるTweetは過去8日分までである。

2.3 データベースの拡張

[2]のシステムは、単語から連想される色彩情報を基に、入力テキストを解析することで、テキストから連想される色彩を推定しており、単語データ・影響度データ・色彩ベクトルデータがプリミティブワード用と未知語用の2種類ずつあるため、計6つのデータベースで構成されている。本研究では[2]のニュース記事と歌詞から得た779語のプリミティブワード以外に、マイクロブログのテキストを用いて新たにプリミティブワードの抽出を行い、被験者実験により各単語の色彩ベクトルを調査した。その結果、新たに148語のプリミティブワードを得た。

2.4 Mecab 辞書の追加

テキストの形態素解析には、Mecab0.996を用いた。マイクロブログ上のテキストはネットスラングや固有名詞が多く、MecabのIPA辞書では適切に形態素解析出来ない。そのため、本研究では、Mecabの辞書にWikipediaのページタイトルとはてなワードを名詞として追加することで、ネットスラングや固有名詞に対しても適切に形態素解析をできるようにした。

連絡先: 坂本真樹, 電気通信大学大学院情報理工学研究科,
東京都調布市調布ヶ丘 1-5-1, Tel: 042-443-5535, FAX:
042-443-5535, sakamoto@hc.uec.ac.jp

3. 結果・考察

本手法では、各テキストから連想される色彩上位 10 色が出力される。色彩候補となる 35 色が、出力された色彩の上位 3 色(以下第 1 色目, 第 2 色目, 第 3 色目)となった回数を地域ごとにそれぞれ集計し, 集計したデータから上位 10 色を降順に並べたものを出現色彩表とした。

出現色彩表中に出現している色彩について, その色彩の出現に影響を与えたテキスト中の単語を集計し, 各地域における出現数上位 10 単語を色彩ごとに表 1 のように, 出現単語表としてまとめた。

表 1: 第 2 色目が **色彩 1** となったテキスト中の単語

Rt = 0.9	東京	外苑前
1 位	人	銀杏
2 位	大阪	並木
3 位	行く	店
4 位	てる	autumn
5 位	店	南青山
6 位	今日	人
7 位	来る	秋
8 位	なる	紅葉
9 位	食べる	ない
10 位	カレー	野球

複数の地域で同じ色彩が出現したからといって, 地域特性が共通しているとはいえないため, それぞれの色彩が各地域それぞれ独自の特性を反映しているかどうかの基準として, 異なり語率を使用した。異なり語率はその単語群の語彙の多様性を示すため, 異なり語率が高い色彩は各地域の独自の地域特性を反映した色彩といえる。逆に異なり語率が低い色彩は, 出現単語に地域差が認められず, 地域間で共通した地域特性を反映しているといえる。今回, 異なり語率の高低を判断する基準値を出現地域数ごとにわけ, 異なり語率が基準値を超えた色彩を各地域の独自の地域特性を反映した色彩, 基準値以下を共通した地域特性を反映した色彩とした。

6 地域の独自の地域特性を反映しているとされた色彩をまとめると, 表 2 のようになった。なお, 第 1 色目, 第 2 色目, 第 3 色目で重複した色彩は別々にカウントした。表 2 において, 外苑前が最も色彩数が多く, 外苑前の地域特性が他の地域と比べて独自性が高いことがわかる。また, 新宿, 渋谷に関して, 色彩数最も少ないことから, 地域特性の独自性が低いといえる。このことから, 渋谷, 新宿の人々の感性には偏りがなく, 様々な感性を持つ人が現れる地域だといえることができる。

また, 表 2 の色彩の出現単語表には各地域それぞれの特徴的な単語が出現していた。例として, 東京, 外苑前の **色彩 1** の出現単語表は表 1 のようになった。表 1 から, 東京では, 大阪やお店, 食事に関する tweet が反映され, 外苑前では, tweet 収集期間が秋季だったこともあり, 紅葉スポットとしても有名な明治神宮の銀杏並木についての tweet や, 秩父宮ラグビー場におけるラグビーの試合, 神宮球場における野球の試合に関する tweet が反映されたことで **色彩 1** が出力されたことがわかる。

また, 地域ごとに, 他地域と共通した地域特性によって同じ色彩を出力した回数を集計した結果, 表 3 のようになった。た

表 2: 6 地域それぞれの独自の地域特性を反映した色彩

横浜	東京	新宿	品川	渋谷	外苑前
8	18	21	13	21	1
18	21	-	21	-	1
21	27	-	27	-	13
29	1	-	-	-	18
-	-	-	-	-	21
-	-	-	-	-	29

だし, 全地域で共通した回数はカウントしていない。表 3 から, 地域特性が共通した回数が最も多かったのが, 横浜と新宿, 横浜と渋谷, 新宿と渋谷の 8 回であった。このことから, この 3 地域間は感性的地域特性が似通っているということが考えられる。[3]によると, 平日, 休日ともに新宿, 渋谷, 横浜の駅利用者の特性は似通っており, 表 3 の結果は妥当だといえる。

表 3: 他地域と共通した地域特性が反映された回数

	横浜	東京	新宿	品川	渋谷	外苑前
横浜	-	6	8	6	8	4
東京	6	-	6	5	5	2
新宿	8	6	-	6	8	4
品川	6	5	6	-	5	3
渋谷	8	5	8	5	-	4
外苑前	4	2	4	3	4	-

しかしながら, 全地域において, 上位色彩には赤や, 白, 黒などの色が集中した。その原因として, それらの色彩の出現単語表内の単語には簡単な補助動詞や状態動詞といったテキストの内容に関わらず出現する簡単な単語が多かったためだと考えられる。

4. おわりに

各地の感性的地域特性を反映した色彩及び, 地域特性の独自性, 地域特性の類似性を示すことができた。しかしながら, 色彩出力に影響を与えた単語に簡単な補助動詞や状態動詞などのテキストの内容に関わらず出現する単語が多かったため, 上位色彩においては地域による差異がみられなかった。そのため, 色彩の細分化やデータベースの改良などの課題が残った。

参考文献

- [1] 齊藤 裕樹・高山 翼・山上 慶・戸辺 義人・鉄谷 信二 (2014), マイクロブログのジオタグと発言コンテキスト解析による行動予測手法, 『情報処理学会論文誌』, 55(2), 773-781.
- [2] 飯場 咲紀・土斐崎 龍一・坂本 真樹 (2013), テキストの感性イメージを反映した色彩・フォント推薦, 『日本バーチャルリアリティ学会論文誌』, 18(3), 217-226.
- [3] 中人 美香・小野 由樹子 (2008). 東京圏における駅を中心とした移動と消費に関する調査研究. 『JR EAST Technical Review』, 24,15-22