

マイクロブログと気象データを利用した眠気の予測

矢野 裕司[†] 加藤 由花[‡] 横井 健[†]

東京都立産業技術高等専門学校 ものづくり工学科[†]

産業技術大学院大学 産業技術研究科[‡]

1 はじめに

近年 Twitter[1]に代表されるマイクロブログの普及により、インターネット上にリアルタイムで人々が思ったことを短い文章で発信するようになった。この発信は Twitter では tweet と呼ばれており、本文でも tweet と呼ぶ。tweet には、ブログや掲示板と比べ個人のリアルタイムな状況が多く見られ、眠いといった状況も発信されている。一方、既研究では疲労から来る眠気は周辺の温熱環境との関係が強い[2]とされているため、眠気が外気の気温とも関係があると考えられる。

そこで、本研究ではマイクロブログにおける発信人数と気象データを利用して、眠気を予測する。

2 提案手法

本研究では、「眠い」というキーワードを含む tweet を発信した人数を眠気の指標とし、重回帰分析により眠気を予測する。重回帰式は(1)式のように表わされる。

$$\hat{y} = \sum_{i=1}^n \alpha_i x_i \quad (1)$$

なお、 \hat{y} は tweet 人数の予測値、 α_i は i 番目の気象データの回帰係数、 x_i は i 番目の気象データ、 n は気象データの数である。

重回帰分析の説明変数となる気象データとして、最高気温・最高気温の前日比・湿度・風速・気圧及びそれぞれの気象データの2乗、3乗、4乗と天気との21個のデータを用意し、重回帰分析の説明変数を、ステップワイズ法を用いて自由度調整済み決定係数が最も高い組み合わせを選択する。tweet の取得は、Twitter 社が提供している Twitter Search API[3]を利用する。評価

には、(2)式で表わされる自由度調整済み決定係数 \hat{R}^2 を用いる。

$$\hat{R}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \frac{n-p-1}{n-1} \quad (2)$$

なお、 \hat{y}_i は観測 i 日目の tweet 人数の予測値、 y_i は観測 i 日目の tweet 人数の実測値、 n は観測日数、 \bar{y} は y の平均値、 p は使用した気象データの数である。

また、説明変数を選択する際に、重回帰分析により求めた重回帰式の妥当性を調べるために t 検定を行う。

3 実験

本節では、分析対象のデータの詳細及び実験結果を述べ、考察を行う。

3.1 実験方法

気象の状況が、「眠い」というキーワードを含む tweet を発信する人数に及ぼす影響を、東京・大阪・札幌の3つの地点における2010年10月23日から12月31日までの70日間の1日毎の気象データと「眠い」というキーワードを含む tweet を発信した人数から分析した。地点は、Twitter Search API の機能である「geocode オプション」により緯度経度及び半径を指定した。一日当たりの「眠い」というキーワードを含む tweet を発信した平均人数は、東京では473人、大阪では129人、札幌では64人であった。

なお、天気は数値データでないため、曇り及び雨をダミー変数で表した。また、数値で表される各気象データは標準化を行った。

3.2 実験結果と考察

変数の選択を行ったところ、どの地点でも最高気温・湿度・最高気温の前日比を説明変数として分析した場合で最も自由度調整済み決定係数が高かった。その分析した結果の、3つの地点における自由度調整済み決定係数を表1に示す。また、分析によって得られた結果の一例として東京での分析結果を表2に示す。

Estimation of Drowsy Level using Micro-Blog and Weather Data

[†]Yuji Yano

[‡]Yuka Kato

[†]Takeru Yokoi

[†]Tokyo Metropolitan College of Industrial Technology

[‡]Advanced Institute of Industrial Technology

表1 各地点の自由度調整済み決定係数

tweet 取得地点	\hat{R}^2
東京	0.578
大阪	0.485
札幌	0.355

表2 東京での分析結果

	係数	標準誤差	t 値
切片	500.99	13.44	37.28
x_1	130.62	13.91	9.39
x_1^2	-48.81	9.20	-5.31
x_2	42.49	10.30	4.12
x_3^3	-0.36	0.15	-2.40

なお、表2の x_1 は最高気温、 x_2 は湿度、 x_3 は最高気温の前日比である。

また、最高気温のみによる tweet 人数の変化を図1、湿度のみによる tweet 人数の変化を図2、最高気温の前日比のみによる tweet 人数の変化を図3に示す。

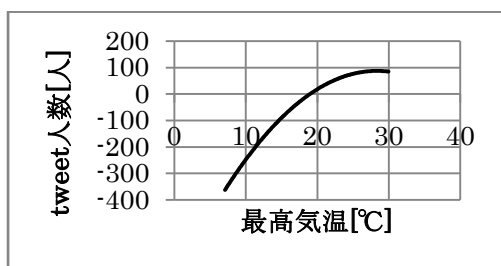


図1 最高気温による tweet 人数の変化

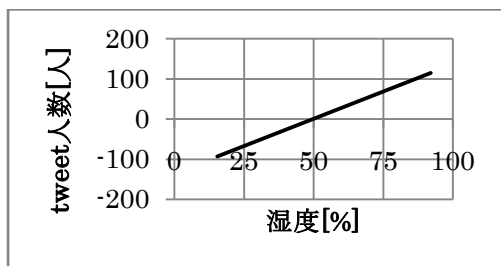


図2 湿度による tweet 人数の変化

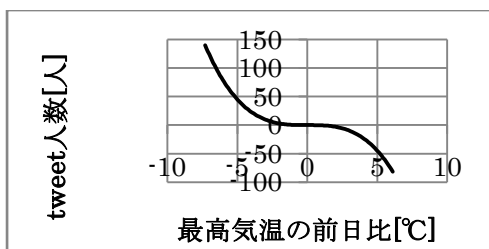


図3 最高気温の前日比による tweet 人数の変化

観測日数は70日で、自由度は69であるため、有意水準1%のt値の境界値は2.649である。よって、切片及び $x_1 \cdot x_1^2 \cdot x_2$ は有意水準1%で有意差があると言える。また、 x_3^3 も有意水準5%のt値の境界値を超えているので、有意水準5%で有意差があると言える。よって、十分に有意であると考えられる。

表1から、3地点の中で東京は自由度調整済み決定係数が最も高く、表2のt値より有意水準1%で有意差があるとわかった。また、大阪でも、自由度調整済み決定係数は0.5近くあり、説明変数により半分近く説明できている。しかし、札幌では0.355とやや少なかった。この理由としては、平均tweet人数は東京が最も多く、大阪では1/4、札幌では1/8程度であるため、データ数の多さが関係していると考えられる。

図1から、最高気温は28.3°Cで眠気を感じる人数が最も多く、87人の増加があった。またその気温から離れると、眠気を感じる人数は少なくなる。図2から、湿度が上昇するに伴って、眠気を感じる人数も多くなる。よって、眠気を感じる人数は比較的天気の良い日に多くなる傾向が見られる。図3から、最高気温の前日比は前日からの温度差が大きいほど重みが増え、眠気を感じる人数に大きく影響する。また、前日よりも最高気温が低下すると眠気を感じる人数が多くなり、上昇すると眠気を感じる人数が少なくなる。

4 まとめ

気象データと「眠い」というキーワードを含むtweetを発信する人数の関係を分析した。その結果、最高気温・最高気温の前日比・湿度が関係しており、眠気にやや大きな影響を与えていることがわかった。また、その分析結果が期間内で有意であることを確かめた。

今後は更に長期的に気象データと「眠い」というキーワードを含むtweetの取得を行い、分析を実施する。

参考文献

- [1]Twitter, <http://twitter.com/>.
- [2]羽田ほか, 「夏季室温緩和設定オフィスにおける温熱環境実測および執務者アンケート調査による知的生産性に関する評価」, 日本建築学会環境系論文集, 74巻, 637号, pp389-396, 2009.
- [3]Twitter Search API, <http://search.twitter.com/>.