E-016

気象情報と Tweet データの統合的分析による体感気温の定量化に関する一考察
Analysis on Quantification of Sensible Temperature by Analyzing Weather Conditions and Tweet Data

馬賀 嵩士 † 三川 健太 ‡ 後藤 正幸 † 吉開 朋弘 §
Takashi Maga Kenta Mikawa Masayuki Goto Tomohiro Yoshikai

1 研究背景・目的

多様な商品を所在地域の異なる多店舗で販売する小売チェーンにおいて，気象条件の変化に伴う需要変動に起因する，在庫過多や廃棄処分が課題となっている．これに対し，蓄積が可能となった多様な大規模データを分析することで，ミクロな観点から需要予測の向上とより細かな在庫管理が期待されている．しかし，特に食料品の需要は絶対的な気象条件よりも，消費者が感じる体感気温に大きな影響を受けると考えられる．

体感気温の定量化に関する試みは数多くなされてきており，湿度や風速といった気象条件に影響を受けやすいといったことがわかっている．代表的な算出式にミスナールの式，リンケの式，NET などがある．しかし，体感気温は湿度や風速，日射量といった気象条件以外にも，着衣量，代謝量のような人体条件の影響も受けるため，その感覚の定量化の方法はさまざまである．これに対し本研究では，人間が感じる体感温度は，Twitter における「暑い」や「寒い」といったつぶやきの形で現れるのではないかという仮定のもと，Tweet データを用いた体感気温の定量化について述べ，その妥当性を検証する．また，体感気温の予測を通して，影響を与えている要因の抽出を図る．

2 事前分析

本研究において用いるデータは，日本語位置情報付き Tweet データ (1/10 サンプリング)，気象データ (アメダス東京地点) の 2 点である．Tweet データは，2012 年 9 月 1 日〜2015 年 9 月 30 日のもので，各 Tweet には，つぶやかれた日時，位置 (緯度と経度) が付与されており，Tweet 総数は 15,495,108 件である．気象データは，期間は Tweet データと同様で，各期 (日付) の平均気温 T [℃]，最低気温 $T^{\min}$ [℃]，最高気温 $T^{\max}$ [℃]，相対湿度 H [%]，風速 W [m/s]，日合計降水量 TRF [mm]，日最深積雪 MSF [cm]，日射時間 ST [時間]，日合計全天日射量 TSR [MJ/m^2] が蓄積されたものである．以降，t 期の気象要素は，各記号に添字 t をつけて表すものとする．ここで，Tweet データを温度感覚の定量化に用いるために，全 Tweet に対して形態素解析を行い，「暑い」という単語を含む Tweet 数の (総 Tweet 数に占める) 割合，「寒い」という単語を含む Tweet 数の割合を日毎に集計した．これらの値と平均気温の推移を図 1 に示す．

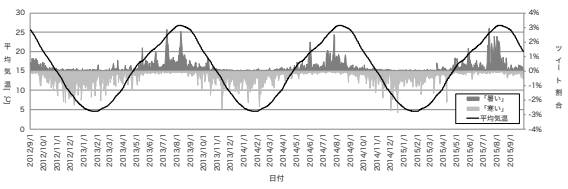

図 1.「暑い/寒い」Tweet の割合と平均気温の推移

図 1 より，「暑い/寒い」Tweet の割合と平均気温の間には一定の相関があり，温度感覚の定量化に用いることの妥当性が明らかとなった．一方で，気温と「暑い/寒い」Tweet の割合との関係に大きくずれが生じている日も存在しており，これらの日では体感気温と平均気温の間に乖離が生じているといえる．次節では，「暑い/寒い」Tweet の割合は体感気温により定まっているのではないかという仮定のもとで，「暑い/寒い」Tweet の割合に基づく体感気温の定義について述べる．また，以降は体感気温と比較する際には，平均気温を実平均気温と表現することとする．

3 Tweet データを用いた体感気温の定義

3.1 概要

t 期の総 Tweet 数に占める「暑い」を含む Tweet の割合を r_t^{hot}，「寒い」を含む Tweet の割合を r_t^{cold}，日平均気温を T_t[℃] とする．このとき，T_t を説明変数，r_t^{hot} を目的変数とした回帰分析を行うことで，偏回帰係数 $\hat{\alpha}_0, \hat{\alpha}_1$ で定められる，その日の気温から「暑い」Tweet 割合を予測する $\hat{r}_t^{hot} = f(T_t|\hat{\alpha}_0, \hat{\alpha}_1)$ なる回帰式が得られる (r_t^{cold} も同様)．一方で，これは同時に，r_t^{hot} からその値に対応する気温を算出する関数 $g(r_t^{hot}|\hat{\alpha}_0, \hat{\alpha}_1)$ が得られたとみなすこともできる．つまり，この関数 g により，r_t^{hot} の値を平均的に生みだし得る気温 S_t [℃] を算出することができる．本研究では，この S_t を t 期の体感気温と定義する．

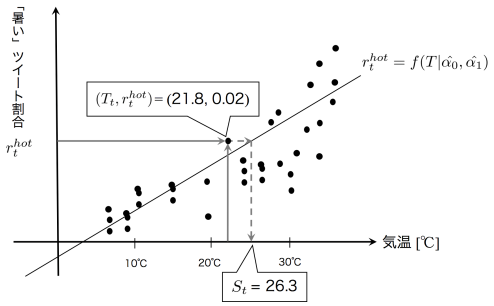

図 2. 体感気温算出の図解

3.2 体感気温の定量化

前節で述べた考えに基づく，ロジスティック回帰分析を用いた体感気温 S_t の算出の流れについて述べる．

Step1) 偏回帰係数の推定

以下の式で定義されるロジスティック回帰分析を行い，偏回帰係数 $\hat{\alpha}_0, \hat{\alpha}_1$ を推定する．

$$\mathrm{logit}(\hat{r}_t^{hot}) = \log\left(\frac{\hat{r}_t^{hot}}{1-\hat{r}_t^{hot}}\right) = \hat{\alpha}_0 + \hat{\alpha}_1 T_t \quad (1)$$

Step2) 変換式の定義

Step1 で得られた回帰パラメータ $\hat{\alpha}_0, \hat{\alpha}_1$ を用いて，以下の式により r_t^{hot} を体感気温 S_t^{hot} に変換する．

† 早稲田大学
‡ 湘南工科大学
§ 一般財団法人 日本気象協会

$$S_t^{hot} = g(r_t^{hot}|\hat{\alpha}_0, \hat{\alpha}_1) \quad (2)$$

$$g(r_t^{hot}|\hat{\alpha}_0, \hat{\alpha}_1) \stackrel{\text{def}}{=} \frac{1}{\hat{\alpha}_1}\left(\log \frac{r_t^{hot}}{e^{\hat{\alpha}_0}(1-r_t^{hot})}\right) \quad (3)$$

Step3) r_t^{cold} に関しても **Step1,2** を行い，S_t^{cold} を算出する．

Step4) 体感気温の算出

以下の式により，体感気温 S_t を算出する．

$$S_t \stackrel{\text{def}}{=} \frac{r_t^{hot}}{r_t^{hot}+r_t^{cold}} S_t^{hot} + \frac{r_t^{cold}}{r_t^{hot}+r_t^{cold}} S_t^{cold} \quad (4)$$

4 体感気温の予測

体感気温を実際の需要予測に用いるためには，数日後の体感気温を知る必要がある．その際，Tweet データを用いることを前提としてしまうと，予測対象日の当日になって Tweet データを取得してからでないと，体感気温を算出できないことになってしまう．そこで，本節では，体感気温に影響を与えている要因の抽出を副次的な目的とした体感気温 S_t の予測について述べる．想定する状況は，想定 **1**. t 期の気象データを入力として S_t を推定，想定 **2**. $(t-1)$ 期の気象データをもとに推定した $\hat{S}_{t-1}$ を入力として S_t を予測，想定 **3**. $(t-1)$ 期の気象データをもとに推定した $\hat{S}_{t-1}$ を用いて予測した $\widetilde{S}_t$ と t 期の気象データを入力として S_t を予測，の 3 パターンとする．想定 1 では重回帰モデルを，想定 2 では重回帰モデルを用いたのち自己回帰モデルを，想定 3 では $\widetilde{S}_t$ の予測までは重回帰モデルと自己回帰モデルを併用し，S_t の予測値算出に再度重回帰モデルを用いるものとする．

4.1 重回帰モデル [1] による予測

t 期の気象要素ベクトルを $\boldsymbol{x}_t \in \mathbb{R}^{d+1}$ とし，$\{(\boldsymbol{x}_t, S_t)\}_{t=1}^N$ を N 個の学習データ集合とする．このとき，$\boldsymbol{x}_t$ を説明変数として S_t を予測する重回帰モデルは以下の式で表される．

$$S_t = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_t + \varepsilon_t \quad (5)$$

ただし，$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)^{\mathrm{T}}$ とし，ε_t は独立に平均 0，分散 σ^2 の正規分布に従う誤差とする．t 期の気象データは $(t-1)$ 期の時点で予報により得られるものとすれば，$(t-1)$ 期の時点で t 期の体感気温を式 (5) で算出することができる．

4.2 自己回帰モデル [2] による予測

時系列の体感気温 $S_1, S_2, \ldots, S_N$ が与えられた場合，この時系列を表現する自己回帰モデルは，

$$S_t = \sum_{i=1}^{p} a_i S_{t-i} + v_t \quad (6)$$

のように表される．p は何期前までの値を考慮するかを決めるパラメータ (次数とよばれる) であり，AIC 基準などに基づいて決定されることが多い．a_i は自己回帰係数であり，v_t は平均 0，分散 σ_v^2 に従う白色雑音である．

5 実験

3 節で定義した体感気温の算出法により，実際の Tweet データと気象データを用いて各期の体感気温の算出を行い，その妥当性を検証する．また，それらの体感気温の予測を通して，影響を与えている要因の抽出を試みる．

体感気温の算出は，2012 年 9 月 1 日〜2015 年 9 月 30 日の全 1125 日間に対して行った．また，予測については，4 節のはじめに述べた 3 パターンの想定について，前半 760 日間 ($N = 760$) の体感気温，および同期間の気象データを学習データ，後半 365 日間 ($N_{test} = 365$) の同データをテストデータとして実験を行った．本実験においては，1 期間は 1 日間を指すものとする．予測精度は平均絶対誤差により評価し，その値が最も小さかったモデルの結果について考察を行う．重回帰モデルに用いる説明変数 $\boldsymbol{x}_t$ は，$\boldsymbol{x}_t = (1, T_t, T_t - T_{t-1}, H_t, W_t, TRF_t, TSR_t, x_{t7}, x_{t8}, x_{t9}, x_{t10})$ であり，$x_{t7}, x_{t8}, x_{t9}, x_{t10}$ は平均気温 T_t について，$T_t \geq 25$，$25 > T_t \geq 20$，$20 > T_t \geq 15$，$15 > T_t \geq 10$ のときそれぞれ 1 をとるダミー変数とする．また，自己回帰モデルの次数 p は 1 とし，適用においては，S_t の階差をとり，対数変換を行ったデータを用いた．想定 3 では，S_t の予測における重回帰に用いる説明変数は，$\widetilde{S}_t, H_t, W_t, TRF_t, TSR_t, x_{t7}, x_{t8}, x_{t9}, x_{t10}$ の 9 変数とした．

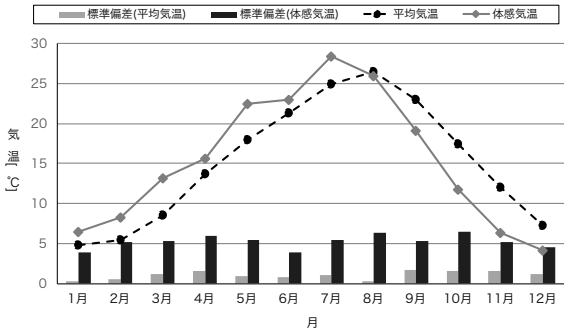

図 3. 実平均気温と体感気温 [℃]

各月における平均気温と体感気温の月平均値の推移，および各月の標準偏差を図 3 に示す．図 3 より，1 月〜7 月は平均気温に比べて体感気温の方が高く，8 月〜12 月においては体感気温の方が低くなっていることがわかる．また，短期間での気温変化が多い 3 月と 5 月，気温が急激に上昇し始める 7 月，および下降し始める 10 月あたりに両者の差が大きくなっており，急激な気温変化の体感気温への影響が示唆される結果となった．また，標準偏差 (図中棒グラフ) に注目すると，特に 2 月，8 月において平均気温と比較して体感気温のばらつきが大きくなっており，この時期は気温以外の要因が体感気温に大きな影響を与えていると考えられる．

次に体感気温の予測の結果について述べる．各モデルの平均絶対誤差は，想定 1: 3.565，想定 2: 4.039，想定 3: 3.504 となった．この結果から，推定値 $\hat{S}_{t-1}$ を用いて時系列予測を行うと精度が低下してしまうこと，当日の気象条件を用いることで予測精度が向上することの 2 点が明らかとなった．ここで，精度が最も良かった想定 3 における標準偏回帰係数を表 1 に示す．

表 **1.** 想定 **3** における標準偏回帰係数

定数項	体感気温予測値 S_t	H	W	TRF	TSR	x_{t7}	x_{t8}	x_{t9}	x_{t10}
15.710	4.307	3.684	-0.330	-0.432	4.030	0.737	-0.114	-1.182	-1.865

表 1 より，その日の体感気温の高さは，前日の体感気温の推定値に基づくその日の体感気温の予測値，相対湿度，日合計全天日射量に大きく影響を受けていることがわかる．また，風速，日合計降水量などが体感気温に負の影響をもたらす結果となっていることから，算出された体感気温の妥当性が検証されたといえる．

6 まとめと今後の課題

本研究では，Tweet データを用いた体感気温の定量化の方法を定義した．また，その体感気温について予測モデルを適用することで，影響を与えている要因の抽出を試みた．今後の課題としては，気象予報の的中率を考慮に入れること，平年比を考慮した気温の考慮方法，気候区分に基づく地域ごとの検証，温度感覚に関わる単語の抽出と利用などが挙げられる．

参考文献

[1] 久保拓弥, “データ解析のための統計モデリング入門”，岩波書店 (2012).
[2] 北川源四郎, “時系列解析入門”，岩波書店 (2005).