

# 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とその需要予測への応用

馬賀 嵩士<sup>†\*a)</sup> 三川 健太<sup>††b)</sup> 後藤 正幸<sup>†c)</sup> 吉開 朋弘<sup>†††d)</sup>

Quantification of Sensible Temperature by Analyzing Meteorological Information and Tweet Data and Its Application to Demand Forecasting

Takashi MAGA<sup>†\*a)</sup>, Kenta MIKAWA<sup>††b)</sup>, Masayuki GOTO<sup>†c)</sup>, and Tomohiro YOSHIKAI<sup>†††d)</sup>

あらまし 近年では、小売店における在庫過多や廃棄処分が課題となっている。この市場における需要予測を考えた場合、気温や天候などの気象条件はその重要な要因であるといえる。その一方で、商品需要は気温などの単純な気象情報ではなく、消費者が感じる体感気温の影響を受けることが想定される。体感気温の定量化に関する研究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。加えて、体感気温はこれらの気象条件以外にも、着衣量、代謝量といった人体条件の影響も受けるといわれており、その感覚の定量化の方法はさまざまである。それに対し本研究では、大規模データである Twitter のデータを学習し、需要予測のための体感気温の定量化を行うモデルを提案するとともに、需要予測への利用の有効性について検討を行う。

キーワード 気象情報, SNS, Twitter, 体感気温, 需要予測

## 1. ま え が き

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、ミクロな観点からの需要予測の精度向上と、より細かな在庫管理が期待されている。本研究では特に、あるカテゴリーの食料品の需要は、客

観的な気温よりも消費者が感じる主観的な体感気温に直接的に大きな影響を受けるのではないかという仮説に着目する。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。この点に着目した代表的な算出式に Missenard の式、*Net Effective Temperature*、不快指数などがある [1]~[3]。加えて、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り、ある作業に特化した形で体感気温の定量化が可能となれば実問題への適用へも繋がると考えられる。しかし、より一般的に“ある気象条件”のもとで消費者全体が感じている体感気温を定量化できれば、小売店の売上予測だけでなく、様々な応用問題に活用できると考えられる。そこで本研究では、Twitter に投稿される Tweet データのうち、温度感覚にかかわるつぶやきを利用することで、人々がその時々を感じている体感気温を定量化することを考える。Twitter は、日本国内のユーザ数が約 3000 万人、1 日の Tweet 数は数億件にも及ぶ大規模 SNS である。数千万人の人々

<sup>†</sup> 早稲田大学創造理工学部経営システム工学科, 東京都  
Department of Industrial and Management and Systems  
Engineering, School of Creative Science and Engineering,  
Waseda University, Shinjyuku-ku, Tokyo, 169-8555 Japan

<sup>††</sup> 湘南工科大学工学部, 藤沢市  
Faculty of Engineering, Shonan Institute of Technology,  
Fujisawa-shi, 251-8511 Japan

<sup>†††</sup> 一般財団法人日本気象協会, 東京都  
Japan Weather Association, Toshima-ku, Tokyo, 170-6055  
Japan

\* 現在, (株) 日立製作所

a) E-mail: kks.w.1214-dax@fuji.waseda.jp

b) E-mail: mikawa@info.shonan-it.ac.jp

c) E-mail: masagoto@waseda.jp

d) E-mail: yoshikai.tomohiro@jwa.or.jp

DOI:10.14923/transinfj.2017JDP7086

の感覚がテキストの形でリアルタイムに蓄積されるこの Tweet データを利用することで、人々が感じている体感気温をより忠実に定量化できるのではないかと考えられる。

そこで本研究では、人間が感じる体感気温は、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という仮定のもと、大規模テキストデータである Twitter のデータから、体感気温の定量化を行うモデルを提案する。ここで、体感気温は人々が感じている気温レベルを意味しており、いわゆる正解値が存在しない。そのため、提案した体感気温の精度を直接的に評価できない。そこで、某小売業における需要予測問題へ適用し、従来の体感気温指標に比べて、提案した体感気温が有効であるか否かを検証する。その際、従来の気象データのみで需要を予測したモデルに比べ、体感気温と需要量の関係という観点から各商品の特性を考察することが可能であることも示す。

本研究で提案する体感気温の定量化手法は、Tweet データを平均気温で説明する回帰モデルを構築し、その逆関数によって本来あるべき体感気温を推測する手法となっている。更に、大手小売チェーンの売上データを用いて、提案モデルにより得られる体感気温と商品需要との関係を分析し、需要予測への適用可能性について検討を行うとともに、体感気温で定義される「暑い日」「寒い日」に売れ行きが伸びる商品カテゴリーの抽出を試みる。その際、小売の現場では Tweet データを収集する手間やコストをかけることが現実ではなく、また未来の需要を予測するために同じ日の Tweet データを事前に得ることは不可能であるということから、構築された体感気温の定量化モデルを用いて気象データのみから将来の体感気温を予測する方法について検討する。その結果、本研究で提案する体感気温の算出方法は、通常気温を用いたときと比べ、特に 9、10 月の気温予測が難しい時期における幾つかの気象感応度の高い商品カテゴリーに対する需要予測精度の観点から優れていることを示す。加えて、消費者の感じる体感気温を考慮して管理を行うべき商品カテゴリーを明らかにし、小売業における気象情報活用の新たな観点を示す。

## 2. 準備

### 2.1 体感気温

体感気温とは、人間の肌を感じる温度の感覚を定量的に表したものである [4]。ほかにも体感温度、温熱

指標、温熱感指標など様々な呼称が存在するが、本論文では体感気温という表現で統一することとする。人間の温度感覚はおおむね、気温の高低に影響を受けることは明らかであるが、気温以外の要因にも影響を受けることが知られており、例えば、風速が 1m/s 上がると体感気温が 1°C 下がるとも言われている [5]。また、湿度が高くなるほど蒸し暑く感じるといったことも広く知られている。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、服装、着衣量、代謝量のような人体条件の影響も受けるため、適用範囲や用途に応じて様々な指標が存在する。

体感気温には、環境側の要因である気温  $T$ 、湿度  $H$ 、風速  $V$ 、放射  $R$  の 4 要素と、人体側の要因である着衣量  $C$ 、代謝量  $M$  の 2 要素が深く関わっていると言われている。これらの要因を考慮した代表的な体感気温の指標を表 1 に示す。Missenard は、気温と湿度を考慮した式

$$MC = T - \frac{1}{2.3} (T - 10) \left( 0.8 - \frac{H}{100} \right) \quad (1)$$

を考案した。ただし、この式の適用範囲は、比較的高い湿度のときに限られる。その後この式は、Gregorczyk により風速の影響を追加した式 (2) に形に改良された。

$$NET = 37 - \frac{37 - T}{0.68 - 0.0014H + \frac{1}{1.76 + 1.4V^{0.75}}} - 0.29T \left( 1 - \frac{H}{100} \right) \quad (2)$$

また、有効温度 ( $ET$ ) とは、気温、湿度、風速の 3 要素で快適さの度合いを表した指標のことである。現在の気温、湿度、風速における温度感覚が、湿度 100%、風速 0.1m/s の条件下でのどの気温に相当するかを

表 1 代表的な体感気温指標一覧 [12]  
Table 1 The list of representative sensible temperature index.

指標	略称	考慮する要因					
		$T$	$H$	$V$	$R$	$C$	$M$
Missenard Climatique	$MC$	○	○				
Net Effective Temperature	$NET$	○	○	○			
有効温度	$ET$	○	○	○			
新有効温度	$ET^*$	○	○	○	○	○	○
標準新有効温度	$SET^*$	○	○	○	○	○	○
作用温度	$OT$	○				○	
風冷指数	$WCI$	○		○			
不快指数	$THI$	○	○				
湿球黒球温度	$WBGT$	○	○	○	○		
予測平均温冷感申告	$PMV$	○	○	○	○	○	○

表したものである．これに加えて放射，着衣量，代謝量を考慮に入れ，基準となる湿度を 50%にしたものが新有効温度 ( $ET^*$ ) である [6]．標準新有効温度 ( $SET^*$ ) [7] では，この新有効温度に，標準環境として風速 0.135m/s，代謝量  $M$ [met]，標準着衣量を与え，異なった条件下での温度感覚の比較を可能としている．また，風冷指数 ( $WCI$ ) [8] は，気温と風速に基づき，

$$WCI = (33 - T)(10.45 + 10V^{0.5} - V) \quad (3)$$

で算出される．極地における寒冷環境の評価に用いられ，冬季の寒さが厳しい北米や北欧で広く利用される．対して不快指数 ( $THI$ ) は，蒸し暑さを数量的に表した指標であり，

$$THI = 0.81T + 0.01H(0.99T - 14.3) + 46.3 \quad (4)$$

で計算される．不快指数が 75%を超えると人口の 1 割が，80%を超えると全人口が不快であると感じると言われている．この指標は，風速や放射の影響が含まれないため，室内用の指標であるといえるが，現在では屋外で使われることもある．湿球黒球温度 ( $WBGT$ ) [9] は，米軍兵士の熱中症予防のために作成された実験式であり，乾球温度，自然湿球温度，黒球温度の三つの温度を計測した環境側の 4 要素を把握することで計算が可能である．予測平均温冷感申告 ( $PMV$ ) [10] は，人体を均一層と仮定した熱収支モデルと被験者実験による温冷感申告を組み合わせて作成される．温度を用いず， $-3 \sim +3$  の 7 段階に分類される．快適方程式に基づいた指標であり，建築や環境工学の分野でよく扱われる．これらの指標は作成時の環境や人体条件に左右される場合が多いため，実際の温熱環境下への適用や拡張には注意が必要である．ここに述べたものをはじめとする指標について，その適用範囲に関する長所・短所を分析した研究も報告されている [11]．

## 2.2 データ概要と基本分析

本研究において用いるデータは，位置情報付き日本語 Tweet データ (1/10 サンプルング)，気象データ (アメダス東京地点)，某小売チェーンのある店舗における商品カテゴリごとの日別売上数量データ，及び気象データ (店舗所在地) の 4 点である．Tweet データは，2012 年 9 月 1 日から 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．この Tweet データは，日本気象協会が Twitter 社から購入した位置情報付きの Tweet データであり，

そちらを正式な契約のもとで本研究の分析に利用した．位置情報は携帯端末から付与された GPS 情報 (ジオタグ) である．ただし，Tweet データには位置情報が付与されているものの，ほとんどの Tweet は関東中心部と関西中心部に極端に集中していたため，本研究では位置情報を用いた細かい分析は断念し，関東中心部の Tweet データのみを用いることとした．気象データの期間は，Tweet データと同様であり，第  $t$  日 (日付) の平均気温  $T_t$  [ $^{\circ}\text{C}$ ]，最低気温  $T_t^{\text{min}}$  [ $^{\circ}\text{C}$ ]，最高気温  $T_t^{\text{max}}$  [ $^{\circ}\text{C}$ ]，相対湿度  $H_t$  [%]，平均風速  $W_t$  [m/s]，日合計降水量  $RF_t$  [mm]，日最深積雪  $SF_t$  [cm]，日照時間  $ST_t$  [時間]，日合計全天日射量  $SR_t$  [ $\text{MJ}/\text{m}^2$ ] を利用する．売上数量データは，当該店舗で販売されている全 153 カテゴリー (中分類) の 2013 年 1 月 1 日から 2014 年 12 月 31 日における日別売上数量が蓄積されたものである．また，このデータには同期間における店舗所在地の気象データが付随しており，平均気温，相対湿度，平均風速，日合計降水量，日合計全天日射量の五つの観測値が含まれている．また，モデル構築を行う対象地域として，Tweet データを関東中心部で絞り込んでいることに合わせ，気象データについても関東中心部である東京の気象データを用いることとする．

ここで「暑い」「寒い」という単語を含む Tweet を温度感覚の定量化に利用することの妥当性を検証するために，全 Tweet データに対して形態素解析 [13] を行い，基本形が「暑い」「寒い」である単語を含む Tweet の総 Tweet 数に占める割合を日ごとに集計した．抽出された「暑い」の Tweet 数は 41,366 件，「寒い」の Tweet 数は 71,044 件であった．これらの Tweet の割合と平均気温の推移を図 1 に示す<sup>(注1)</sup>．図 1 より，「暑

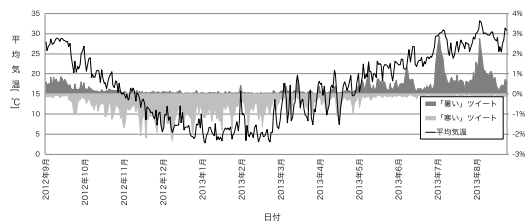


図 1 「暑い/寒い」 Tweet の割合と平均気温の推移  
Fig.1 Transitions of hot/cold tweets rate and average temperature.

(注1) : Tweet の絶対数は Twitter の普及に伴うユーザ数増加などにより増加傾向にあるため，Tweet 数そのものではなく，全 Tweet 数に対する割合を用いている．

い/寒い」を含む Tweet の割合と平均気温の間には一定の相関があり、これらの値を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。なお、平均気温と「暑い」の Tweet 割合との相関係数は 0.656、「寒い」の Tweet 割合との相関係数は  $-0.680$  であった。若干、「寒い」の Tweet 割合の方が相関係数の絶対値が大きいが、ほぼ同等の相関であることがわかる。一方で、平均気温と「暑い/寒い」を含む Tweet の割合の挙動に大きくずれが生じている時期も存在しており、これらの時期においては体感気温と平均気温の間に乖離が生じている可能性があることがわかる。

次に、「暑い/寒い」を含む Tweet の割合と商品需要との関係を確認するため、平均気温と「暑い/寒い」を含む Tweet の割合、及び平均気温と商品需要の関係を図 2 に示す。ただし、図 2 における“降温期”は 2-8 月を指し、“昇温期”は 3-8 月を示すものとする。なお、本研究では気象感応度が特に高い商品カテゴリーを分析対象とするものとし、暑さに影響を受けやすい商品カテゴリーとして「飲料」を、寒さに影響を受けやすい商品カテゴリーとして「菌茸・鍋」を取り上げた。

なお、これらの図の「暑い/寒い」を含む Tweet の割合と平均気温の相関係数は、先に示したように、それぞれ 0.656 と  $-0.680$  であり、「飲料」の売上数量と平均気温の相関係数は 0.633、「菌茸・鍋」の売上数量と平均気温の相関係数は  $-0.468$  であった。

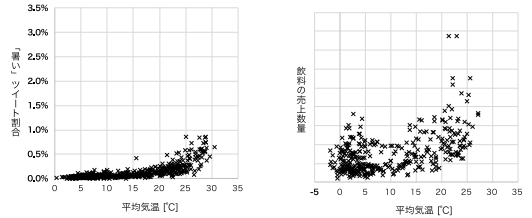
ここで、Tweet データには「暑くない」、「寒くない」などの否定表現や「暑いかな?」、「寒くない?」といった疑問形、「ライブが暑い」という漢字を「ライブが暑い」と誤って入力した文章などが含まれることが考えられ、様々なノイズが混入する可能性がある。しかしながら、本研究では、

- これらのノイズを除去しなくても、平均気温や需要量との一定の相関関係が見られており、妥当性が認められること

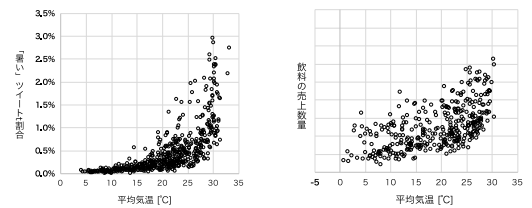
- ノイズは時間軸に対して独立にランダムに発生すると仮定すれば、後の章で導入する回帰モデル構築の際に最小 2 乗法によって除去される

という理由から、これらのノイズの混入は無視して分析を進めることとした。これらのノイズの割合は単純な「暑い」「寒い」という Tweet と比べると、実際に非常に小さく、ネガボジ分析などによる否定後の除去などの影響は限定的であると考えられる。

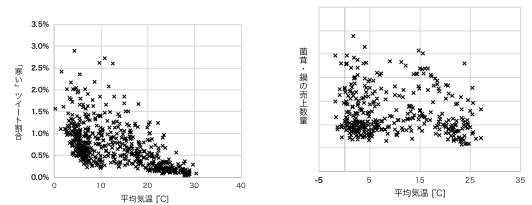
図 2-(a) より、「暑い」を含む Tweet の割合は、実平均気温に対して曲線的に増加しており、飲料の売上



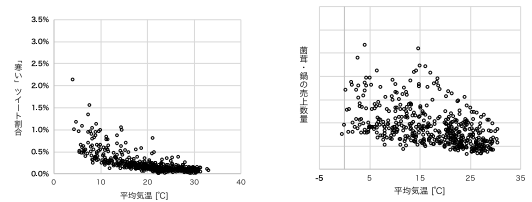
(a) 「暑い」 Tweet 割合と飲料の売上数量 (降温期)



(b) 「暑い」 Tweet 割合と飲料の売上数量 (昇温期)



(c) 「寒い」 Tweet 割合と菌茸・鍋の売上数量 (降温期)



(d) 「寒い」 Tweet 割合と菌茸・鍋の売上数量 (昇温期)

図 2 平均気温と Tweet 割合、売上数量の関係の比較

Fig. 2 Comparison of relationship between average temperature and hot/cold tweets rates, and sales volume.

数量も似た傾向を示していることがわかる。また、「飲料」の売上数量では、昇温期、降温期ともに平均気温が  $15^{\circ}\text{C}$  前後において小さな山ができており、この付近の体感の違いが飲料の需要の変動に影響を及ぼしていると考えられる。次に、図 2-(b) を見ると、「寒い」を含む Tweet の割合が平均気温に対して曲線的に減少しているが、「暑い」 Tweet の場合よりも勾配は緩やかであり、「菌茸・鍋」の売上数量もそれと似た傾向となっていることがわかる。このように、気温に応じた

暑さと寒さの感じ方の傾向には違いが見られることから、本研究においては両者を独立に扱い、暑さ、寒さそれぞれに基づく体感気温を定量化するものとする。

### 2.3 体感気温算出モデルの構築と予測に用いるデータ

前節で示したように、「暑い/寒い」を含む Tweet 割合を用いて体感気温の算出式を構築すること自体には一定の合理性があると考えられる。一方で、このようにして得られた体感気温の算出式を実際に現場で用いて、需要予測などの日々のオペレーションに活用しようとする場合には、そのような利用場面の事情を考慮する必要がある。すなわち、

(1) モデル構築フェーズ：過去の Tweet データと気象データから体感気温の算出式を推定する段階

(2) 体感気温予測フェーズ：得られた体感気温算出式を用いて、体感気温（または需要）の予測を行う段階

を区別して議論を行う必要がある。このうち、(2)の予測フェーズにおいては、実際の Tweet 割合の値を、そのまま体感気温の算出における入力値として用いることは、以下の3点の理由から実用上好ましくないと考えられる。

- 体感気温の算出に Tweet 割合を用いる形にしまうと、未来の体感気温を予測する際に未来の Tweet 割合を予測する必要が出てしまい、現実的に困難である。
- 小売の現場において、Tweet データを蓄積する手間やコストをかけることは現実的ではない<sup>(注2)</sup>。
- Tweet データの性質上、Tweet 割合の値には一定のノイズが含まれていると考えられるため、そのままの値を用いてしまうとノイズの影響を大きく受ける可能性がある。

そこで本研究では、過去の Tweet データと気象データから構築された体感気温の算出式を用いて、体感気温の予測に利用する際には、気象データのみを用いた予測方法を検討する。これにより、小売の現場が Tweet データそのものを用いることなく、気象要素で説明可能な範囲での体感気温の予測が可能となる。気象データに関しては、現在はかなりの精度で翌日や

翌々日の気象予測が可能となっており、これを入力とする算出モデルであれば、比較的容易に近い未来の体感気温を予測することができる。これは、体感気温の予測値を実際に活用する小売業などの立場に立てば、大きなメリットとなると考えられる。

以上のことから、本研究では、3.、4.の体感気温の算出においては Tweet データ及び気象データ（アマダス東京地点）を、5.、6.においては、日別売上数量データ及び気象データ（店舗所在地域）を用いるものとする。

## 3. Tweet データを用いた体感気温の定義

本章では、過去の Tweet データと気象データを用い、体感気温の算出式を構築する方法を提案する。まず、その着想について概要を示した後で、体感気温の定量化の具体的な方法を示す。

### 3.1 着想

Twitter を利用しているユーザは、その日に感じた暑さ、寒さに基づいて、「今日はかなり暑い」、「寒すぎて布団から出られない」、のように、「暑い」や「寒い」といった単語を含む形で自身の感じている温度感覚をつぶやいていると考えられる。また、暑さ、寒さの程度に比例して、暑い、寒いと感じるユーザの数は増加し、このような Tweet の数も増加すると考えられるため、これらの Tweet の数が暑さ、寒さの程度を示していると考えerことは妥当であるといえる。よって、これらの値と実平均気温との関係をモデル化することで、体感気温の定量化が可能となるのではないかと考えられる。ただし、ここでは「暑い/寒い」を含む Tweet は絶対数ではなく、その日の総 Tweet 数に占めるそれぞれの Tweet の割合を用いるものとする。いま、第  $t$  日の総 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  $r_t^{cold}$ 、実平均気温を  $T_t$  [°C] とする。本研究では、「暑い/寒い」の Tweet 割合と実平均気温との関係の定量化に回帰モデル [14] を用いるものとする。具体的には、 $T_t$  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、その日の気温から「暑い」の 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。一方で、この逆関数  $f^{-1}$  は、「暑い」 Tweet 割合からその値に対応する気温を算出する関数  $g(r_t^{hot} | \hat{\alpha}_0, \hat{\alpha}_1) = f^{-1}(r_t^{hot} | \hat{\alpha}_0, \hat{\alpha}_1)$  であるとみなすこともできる。すなわち、この関数  $g$  により、「暑い」 Tweet 割合の値を平均的に生みだし得る気温

(注2)：実際に、体感気温を算出しようとする際に、利用する小売業者等が当日の Tweet データを用意することは、技術的な観点からかなり無理がある。また、第三者がデータを用意する場合においても、当日のデータを即日で購入したり、あるいはクロージングして時間をかけてデータを収集しなければならず、体感気温の算出のためのコストがかかってしまう。

$S_t^{hot}$  [°C] を推定することができる．本研究では，この  $S_t^{hot}$  を暑さに基づく第  $t$  日の体感気温と定義する<sup>(注3)</sup>．

ここで，上記の回帰における残差  $\delta_t = r_t^{hot} - \hat{r}_t^{hot}$  は，一定の誤差を含みながら，湿度や風速といった気温以外の気象要素の影響により生じていると仮定し，この  $\delta_t$  をそれらの気象要素から何らかのモデルで回帰することを考える．回帰による推定値を  $\hat{\delta}_t = \delta_t - \varepsilon$  とすると ( $\varepsilon$  は残差)， $\tilde{r}_t^{hot} = \hat{r}_t^{hot} + \hat{\delta}_t$  は気象要素の影響のみを考慮した場合に想定される第  $t$  日の「暑い」Tweet 割合とみなすことができる．この  $\tilde{r}_t^{hot}$  を先に述べた関数  $g$  に入力することで，気象要素の影響のみを考慮した場合に想定される  $S_t^{hot}$  の値を算出することができる． $\delta_t$  から  $S_t^{hot}$  を算出することも可能だが，その場合， $S_t^{hot}$  の算出に  $r_t^{hot}$  の値が必要となるため，Tweet データを蓄積する手間を考えると実用上好ましくない．加えて，Tweet データの性質上， $r_t^{hot}$  の値には一定のノイズが含まれると考えられるため， $\delta_t$  の値を直接  $S_t^{hot}$  の算出に用いるのは適切ではないといえる．そこで，気象要素から  $\delta_t$  を推定し，そのときの推定誤差は気象要素以外の要因に起因するノイズであると考えることで，この問題を解消する狙いもある． $r_t^{cold}$  についても同様の流れを踏むことで，寒さに基づく第  $t$  日の体感気温  $S_t^{cold}$  を算出することができる．また，これ以降， $S_t^{hot}$  と  $S_t^{cold}$  を合わせて体感気温と表現することとする． $r_t^{hot}$  による  $S_t^{hot}$  の算出のイメージを図 3 に示す<sup>(注4)</sup>．

### 3.2 体感気温の定量化

以下では，具体的な  $S_t^{hot}$ ,  $S_t^{cold}$  の算出方法について述べる．まず，平均気温と Tweet 割合との関係を定式化するため，式 (5), (6) のとおり回帰式を定義する．

$$R_t^{hot} = \alpha_0^h + \alpha_1^h T_t + \varepsilon_t \quad (5)$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_h^2) \quad (6)$$

(注3)：本論文では，この後で示すように「暑い」の Tweet データと「寒い」の Tweet データから定量化される体感気温の算出モデルを別々に構築している．これは「暑い」と「寒い」のどちらかに応答する商品需要があるため，別々に指標化しておいた方が後の使い勝手が良いという実務上の理由による．また，体感気温の推定を回帰モデルの逆関数で推定するという方法は，説明変数が二つ以上の重回帰モデルでは説明変数と目的変数の関数形が 1 対 1 対応していないため，適用が難しい．本研究では，気象感応度の高い商品の需要予測に役立つような体感気温の定量化を一義的な目的としており，そのための方法として望ましい方法を検討することに主眼を置いている．

(注4)：このように，Tweet の割合を平均気温で説明する回帰モデルを構築し，その残差について他の気象データから予測できる効果部分を評価したのちに，回帰の逆関数を用いて体感気温を推測するというアプローチはあまり見られない方法である．

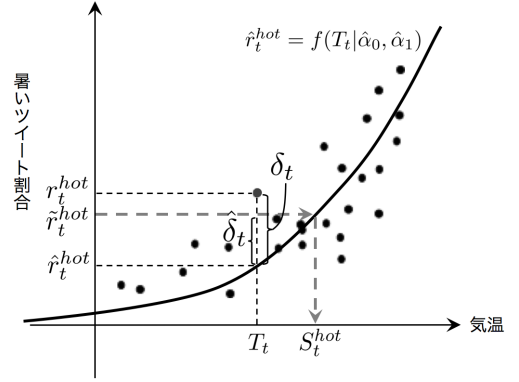


図 3 体感気温算出のイメージ  
Fig. 3 Schematic illustration of quantifying the sensible temperature.

ただし，説明変数を  $T_t$ ，目的変数を  $R_t^{hot} = \log(r_t^{hot}/(1 - r_t^{hot}))$  とする．また，誤差  $\varepsilon_t$  は独立に平均 0，分散  $\sigma_h^2$  の正規分布に従うものとする．次に，平均気温以外の気象要素からなる気象要素ベクトルを  $\mathbf{w}_t$  として，この回帰における残差  $\Delta_t^{hot} = R_t^{hot} - (\hat{\alpha}_0^h + \hat{\alpha}_1^h T_t)$  を目的変数とし， $\mathbf{w}_t$  を説明変数とした回帰モデルを構築する．ただし，各気象要素の体感気温への影響の度合いは季節ごとに異なるという仮定から，回帰モデルは季節ごとに構築するものとする． $s_t$  を第  $t$  日が春 (3-5 月) であるとき 1，夏 (6-8 月) であるとき 2，秋 (9-11 月) であるとき 3，冬 (12-2 月) であるとき 4 をとる季節変数とすると，季節  $k$  における  $\Delta_t^{hot}$  の回帰式は以下のように表せる．

$$\hat{\Delta}_t^{hot} = F_k^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) \quad (7)$$

$$\forall t \in \mathcal{S}_k := \{t | s_t = k\} \quad (8)$$

ここで， $F(\cdot | \boldsymbol{\theta})$  は，パラメータを  $\boldsymbol{\theta}$  とする任意の回帰モデルであるとする．更に，ここまでの手順を  $R_t^{cold} = \log(r_t^{cold}/(1 - r_t^{cold}))$  についても同様に行い， $\hat{\alpha}_0^c, \hat{\alpha}_1^c$  を推定， $F_k^{cold}$  を学習する．そして，得られた  $\hat{\alpha}_1^h, \hat{\alpha}_1^c, F_k^{hot}, F_k^{cold}$  を用いて，式 (5) の逆関数に対応する以下の式 (9), (10) により  $S_t^{hot}, S_t^{cold}$  を算出する．

$$\begin{aligned} S_t^{hot} &= \frac{1}{\hat{\alpha}_1^h} \left( (\hat{\alpha}_0^h + \hat{\alpha}_1^h T_t) + F_{s_t}^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) - \hat{\alpha}_0^h \right) \\ &= \frac{1}{\hat{\alpha}_1^h} \left( \hat{\alpha}_1^h T_t + F_{s_t}^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) \right) \\ &= T_t + \frac{1}{\hat{\alpha}_1^h} F_{s_t}^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) \end{aligned} \quad (9)$$

$$S_t^{cold} = T_t + \frac{1}{\hat{\alpha}_1^c} F_{st}^{cold} (\mathbf{w}_t | \theta_k^{cold}) \quad (10)$$

体感気温の算出アルゴリズムを以下に示す.

[体感気温の算出アルゴリズム]

Step1) Tweet 割合の回帰

Step1-1) 式 (5), (6) で定義される回帰分析を行い, 偏回帰係数  $\alpha_0^h, \alpha_1^h$  を推定する.

Step2-1) 同様に,  $R_t^{cold}$  についても  $\alpha_0^c, \alpha_1^c$  を推定する.

Step2) 残差回帰

Step2-1) 各季節 ( $k = 1 \sim 4$ ) について,  $F_k^{hot}$  を学習する.

Step2-2) 同様に,  $R_t^{cold}$  についても  $F_k^{cold}$  を学習する.

Step3) 体感気温の算出

式 (9), (10) により  $S_t^{hot}, S_t^{cold}$  を算出する. □

#### 4. 体感気温の算出

本章では, 前節で定義した体感気温の算出方法により, 実際の Tweet データと気象データ (アメダス東京地点) を用いて体感気温の算出モデルを構築し, それを用いて体感気温の算出を行い, 実平均気温との差異などの結果について考察を与える (注5).

##### 4.1 概要

対象期間は, 2012年9月1日から2015年9月30日であり, 「暑い/寒い」の Tweet 割合に対して, 対移動平均比率法 [15] を適用することで曜日効果の除去を行った. また, 気象要素ベクトルは, 気象データ (店舗所在地) に共通する気象要素のみを含むよう  $\mathbf{w}_t = (T_t - T_{t-1}, H_t, W_t, RF_t, SR_t)^T$  とし, 回帰モデル  $F$  には Random Forest [16] を用いるものとする. Random Forest におけるパラメータは Grid Search により決定した.  $\mathbf{w}_t$  の選定に際しては, 以下の基準で行うものとした.

(1) 一般に気象分野において重要と考えられている気象データに含まれる変数を候補とする.

(2) 上記の候補変数の中に, 相関係数が極めて高い変数があった場合には, 重要度の高い一方の変数のみ選択 (選択には気象分野の専門家による判断を加える).

(3)  $\mathbf{w}_t$  の定義より, 気温は除外.

(4) 地域によって極端に数値が異なる変数は除外.

(5) 体感気温に関する既存研究から有効であると示されている変数があれば追加.

#### 4.2 結果

期間内の平均気温  $T_t$  と暑さに基づく体感気温  $S_t^{hot}$ , 及び寒さに基づく体感気温  $S_t^{cold}$  の推移を図 4 に示す (注6). 図 4 より,  $S_t^{hot}, S_t^{cold}$  どちらも, 大まかな傾向は  $T_t$  に類似しているが, ばらつきは  $T_t$  よりもかなり大きくなっていることがわかる. また,  $S_t^{hot}$  と  $S_t^{cold}$  では  $S_t^{cold}$  の方がばらつきが大きくなっており, 9月~12月においてその傾向が顕著に見られることがわかる.

各月における実平均気温と体感気温の月平均値の推移, 及び各月の標準偏差を図 5 に示す. ただし, 図中の  $\mu_T, \mu_{Sh}, \mu_{Sc}$  はそれぞれ  $T_t, S_t^{hot}, S_t^{cold}$  の各月の平均値を,  $\sigma_T, \sigma_{Sh}, \sigma_{Sc}$  は同指標の各月の標準偏差を指すものとする. 図 5 より, 1月から7月は実平均気温よりも体感気温の方が高く, 8月から12月は体感気温の方が低くなっていることがわかる. また,  $S_t^{hot}$  に比

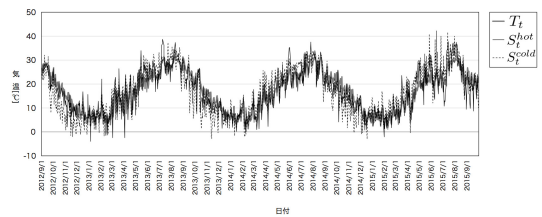


図 4 実平均気温と体感気温の推移  
Fig. 4 Transitions of average temperature and sensible temperature.

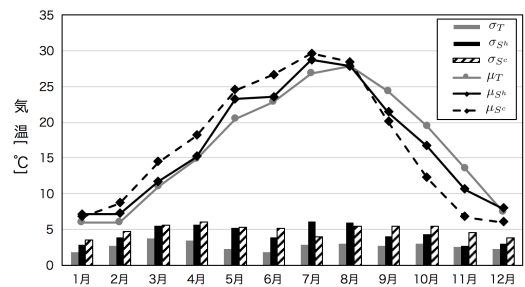


図 5 月ごとの実平均気温と体感気温の比較  
Fig. 5 Comparison between average temperature and sensible temperature.

(注5): 本章における体感気温の算出は, 日次の体感気温を, その日の Tweet データと気象データを用いて推定しており, 過去の状態について遡って体感気温を推測している.

(注6): Missenard の式も同様に平均気温と湿度の関係を用いた体感気温の算出を行っている. 本来であれば図 4 に記載すべきであるが, 平均気温と同様の傾向を示しており, 見やすさのため記載を省略するものとした.

べて  $S_t^{cold}$  の方が体感気温との差が大きくなる傾向にあることがわかる。更に、短期間での気温変化が多い5月、及び下降し始める10月、11月頃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。この差を季節単位で見ると、実平均気温が下がり始める秋(9~11月)には平均して両者の差が大きく、低い気温で安定している冬(12~2月)には平均して両者の差が小さいことがわかる。このことから、人間は「暑い→寒い」の変化に特に敏感であること、冬の時期は低い気温に慣れ、寒さの感じ方があまり変動しないことが考えられる。また、標準偏差(図中棒グラフ)に注目すると、平均気温と比較して、7月~8月において  $S_t^{hot}$  の、9月~2月において  $S_t^{cold}$  のばらつきが大きくなっていることがわかる。これらは、それぞれ「暑い」、「寒い」を含む Tweet が多くなりやすい時期であり、これらの変動に大きく影響を受けたためではないかと考えられる。

次に、 $F_k^{hot}, F_k^{cold}$  の学習の過程で得られた各特徴量の重要度を表2に示す。Random Forest では、学習の過程で特徴量の重要度を推定することができる。この重要度が高い特徴量ほど回帰に有効な特徴量と言えるため、この値を比較することで、体感気温にどの気象要素が大きく影響を与えているかを明らかにすることができる。表2の結果から、春( $k=1$ )の暑さには、他の季節と比較して、日合計全日射量、平均風速が体感気温に大きな影響を及ぼしており、強い風が体感気温を左右すると考えることができる。一方、この時期の寒さは気温の前日差、日合計全日射量に大きく影響を受けており、前日の気温と比較したときの落差が寒さの程度を決める大きな要因になっていると考えることができる。夏( $k=2$ )の暑さには日合計全日射量、相対湿度が大きな影響を与えており、

日差しが強く蒸し暑い夏の特徴と一致している。一方で、この時期に寒いと感じる要因には、気温の前日差が関係していると考えられる。秋( $k=3$ )の寒さには相対湿度、平均風速の影響が大きく、図5から明らかのように、この時期は体感気温が実平均気温に比べて低くなっていることから、湿度の低下に伴う乾燥と強い風が体感気温の低下を引き起こしていると考えられる。一方で、この時期に日差しが強く、前日よりも気温が高くなると暑いと感じる傾向にあることがわかる。冬( $k=4$ )は暑さ、寒さともに特に日合計全日射量の影響が大きく、冬の温度感覚は主に日射量に影響を受けていると考えることができる。

本研究及びこれまでに提案されてきた体感温度・温熱指標のどちらにおいても、相対湿度や風速、日射量などは効果のある要素として取り入れられており、一定の整合性のある結果が得られたと言える。加えて、従来の指標では考慮されていなかった「気温の前日差」が体感気温に比較的大きな影響を与えていることが上記の結果から明らかとなった。また、これまでは、気象要素と体感気温の間に単一の構造を仮定したものがほとんどであったが、本研究の結果から季節ごとに体感気温に影響を与える気象要素、及びその程度は異なるということが明らかとなった。以上から、体感気温を気象要素を用いて厳密に定量化するためには、季節ごとに層別して、モデルを構築することが望ましいと考えられる。

## 5. 体感気温を用いた需要予測

本章では、大手小売チェーンの売上数量データを用いて、本研究で提案したモデルで得られる体感気温が、実平均気温や従来研究で提案されている他の温熱指標よりも需要変動を説明できているかを検証する。

ここでは、同チェーン店で販売されている全商品カテゴリーのうち気象感応度が高く、売上数量が比較的多い、「菌茸・鍋」、「こんにやく」、「練り製品」、「飲料」、「乳飲料・清涼飲料」の5カテゴリーを対象として体感気温を用いた需要予測を行う。そして、実平均気温及び従来の体感温度・温熱指標を用いて需要予測を行った場合の結果との比較から体感気温の需要予測への利用に関する有用性を検証する。

### 5.1 実験条件

用いるデータは、2013年1月1日から2014年12月31日における、大手小売チェーンのある店舗の日ごとの各商品カテゴリーの売上数量データ、及び店舗の

表2 各気象要素の重要度の算出結果

Table 2 Result of importance of each meteorological element.

		$k=1$	$k=2$	$k=3$	$k=4$
$F_k^{hot}$	気温前日差	0.165	0.176	<b>0.217</b>	<b>0.231</b>
	相対湿度	0.197	<b>0.201</b>	0.202	0.208
	平均風速	<b>0.207</b>	0.183	0.175	0.188
	日合計降水量	0.153	0.116	0.172	0.103
	日合計全日射量	<b>0.278</b>	<b>0.324</b>	<b>0.234</b>	<b>0.271</b>
$F_k^{cold}$	気温前日差	<b>0.265</b>	<b>0.264</b>	0.194	<b>0.227</b>
	相対湿度	0.216	0.180	<b>0.296</b>	0.204
	平均風速	0.203	0.197	<b>0.210</b>	0.218
	日合計降水量	0.091	0.114	0.102	0.031
	日合計全日射量	<b>0.224</b>	<b>0.245</b>	0.198	<b>0.320</b>



所在地域の気象データである。既に述べているように、需要予測では、一般に予測段階で Tweet データを得ることが難しい。そこで、同地域の体感気温は、4. で得られたパラメータ及び回帰モデルを式 (9), (10) に代入することで得られる体感気温の算出式に、同地域の気象データを入力することで得られる予測値を用いるものとする<sup>(注7)</sup>。対象期間のうち、前半 365 日間の体感気温、及び同期間の対象商品カテゴリーの売上数量データを学習データ、後半 365 日間の同データをテストデータとして実験を行った。実験対象とする指標は、平均気温  $T_t$ 、*Net Effective Temperature* (*NET*)、不快指数 (*THI*)、寒冷指数 (*WCI*)、 $S_t^{hot}$ 、 $S_t^{cold}$  の六つを用いた。前述のとおり、本研究の目的は体感気温の定量化であるが、参考として、平均気温  $T_t$  に日射量を指標としたもの ( $T_t$  + 日射と略記) も比較することとした。予測には重回帰モデルを用いることとし、用いる説明変数は、指標の値、指標の値の前日差、前々日差、曜日変数とする。ここで、曜日変数とは、金曜日を基準とする六つのダミー変数からなるものとし、それぞれ、土曜日、日曜日、月曜日、火曜日、水曜日、木曜日のとき 1、それ以外のとき 0 をとるものとする。精度評価については、式 (11) で定義される平均絶対誤差率を用いるものとする。ただし、 $N_t$  は第  $t$  日の売上数量、 $\hat{N}_t$  は  $N_t$  の予測値、 $D$  を本指標の算出対象日数とする。

$$\text{平均絶対誤差率 [\%]} = \frac{\sum_{t=1}^D |\hat{N}_t - N_t|}{D} \times 100 \quad (11)$$

### 5.2 実験結果

まず、全期間に対する平均絶対誤差率の結果を表 3 に示す<sup>(注8)</sup>。表 3 より、どの商品カテゴリーにおいても、本研究で提案した体感気温のいずれかが、平均気温及び従来の三つの温熱指標よりも精度の面で優れて

(注7)：これは、気象予測の分野では平均気温や湿度についてはある程度、高精度な予測が可能となってきているという現状を踏まえた設定である。なお、需要予測のために、予測対象日の前日や前々日の気象データを入力することで、予測対象日の体感気温とそれに伴う商品需要を予測するというモデルなど、幾つかのバリエーションについても実験して評価を行ったが、結果はおおむね同様であった。

(注8)：本研究では、式 (11) で示した平均絶対誤差率の他に、

$$\frac{\sum_{t=1}^D |\hat{N}_t - N_t|}{\sum_{t=1}^D N_t} \quad (12)$$

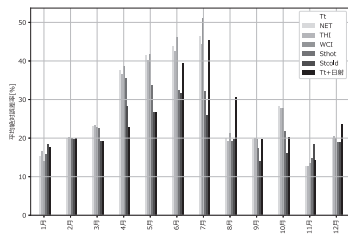
による値の算出も行っている。しかしながら、両者にはほとんど違いが見られなかったことから、式 (11) に絞って説明を与える。

表 3 商品カテゴリーごとの平均絶対誤差率 [%]  
Table 3 Mean absolute percentage error of each item category.

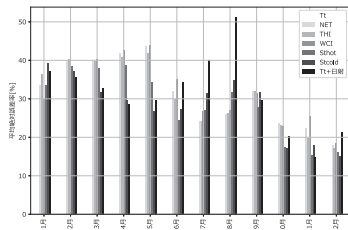
	$T_t$	$T_t$ + 日射	温熱指標			体感気温	
			<i>NET</i>	<i>THI</i>	<i>WCI</i>	$S_t^{hot}$	$S_t^{cold}$
菌茸・鍋	26.9	25.0	27.4	27.0	28.1	23.7	21.4
こんにゃく	30.8	31.3	31.3	30.9	31.8	28.5	28.3
練り製品	21.4	21.1	21.4	21.4	21.6	20.6	20.2
飲料	17.3	15.1	18.0	17.4	19.1	15.1	16.2
乳飲料・清涼飲料	12.7	11.7	12.7	12.7	13.2	12.4	12.9

いることがわかる。更に、 $T_t$  + 日射と比較しても、「乳飲料・清涼飲料」以外の商品カテゴリーでは提案した体感気温が優れていることがわかる。また、「菌茸・鍋」、「こんにゃく」、「練り製品」といった寒さに影響を受けやすいと考えられる商品カテゴリーは  $S_t^{cold}$  による推定精度が良く、「飲料」、「乳飲料・清涼飲料」といった暑さに影響を受けやすいと考えられる商品カテゴリーは  $S_t^{hot}$  による推定精度が良いことがわかる。このことから、 $S_t^{cold}$  は寒さの、 $S_t^{hot}$  は暑さの変動をよく説明できているといえる。従来研究で提案されている温熱指標は、どの商品カテゴリーにおいても平均気温を用いた場合と同程度かそれよりも劣る精度となっており、これらの指標を気象感応度の高い商品の需要予測に用いることは適切ではないことがわかった。提案モデルで算出される  $S_t^{hot}$ 、及び  $S_t^{cold}$  はそれぞれ使い分けることで、更に効果的になる可能性がある。例えば、前者は暑さに影響を受けやすいと考えられる商品カテゴリーの 9~11 月の需要予測を行う際に、後者は寒さに影響を受けやすいと考えられる商品カテゴリーの 4~7 月の需要予測を行う際に用いることが有効であると言える。このように、春季と秋季という気温の上下が比較的大きな季節に使用すべき基準が明確となっている点は本提案モデルの利点であると言える。

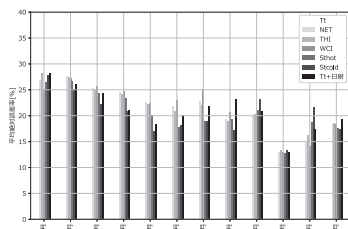
次に、 $S_t^{hot}$  と  $S_t^{cold}$  がどの商品カテゴリーのどの時期の需要予測に特に有効であるかを明らかにするため、各商品カテゴリーの月ごとの平均絶対誤差率を図 6 に示す。まず、図 6-(a)-(c) の「菌茸・鍋」、「こんにゃく」、「練り製品」の 3 商品カテゴリーについて見てみると、どの商品カテゴリーにおいても本研究で提案した体感気温は優れた精度を示しており、特に 5 月~7 月においてこの傾向が顕著に見られる。また、 $S_t^{hot}$  と  $S_t^{cold}$  を比較すると、 $S_t^{cold}$  の方が前述の期間における精度が優れており、 $S_t^{hot}$  よりもこの期間の寒さの変動を説明できているという結果となった。次に、平均気温と従来の温熱指標 (*NET*, *THI*, *WCI*) を比較して



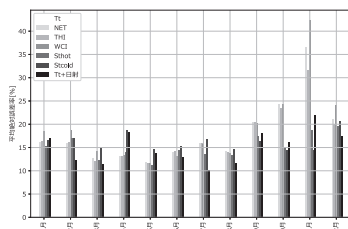
(a) 菌茸・鍋



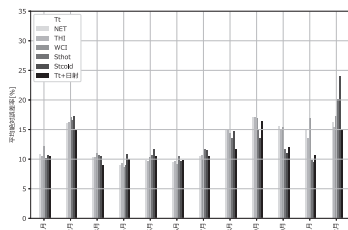
(b) こんにゃく



(c) 練り製品



(d) 飲料



(e) 乳飲料・清涼飲料

図6 月ごとの平均絶対誤差率の比較

Fig. 6 Comparison between mean absolute percentage errors per month.

みると、平均気温と  $NET$ ,  $THI$  の間に違いはあまり見られず、 $WCI$  は、1月、2月あたりでは平均気温よりも良い精度を示すという結果となっている。これは、 $WCI$  は寒さが厳しい時期・地域を想定した指標となっているため、1月、2月がそのような特徴に近い期間として当てはまったのではないかと考えられる。次に、図6-(d)-(e)の「飲料」、「乳飲料・清涼飲料」の2商品カテゴリーを見ると、どちらの商品においても、特に9月～11月において本研究で提案した体感気温が優れた精度を示しており、他の指標よりもこの時期の需要変動を説明できていることがわかった。一方、 $S_t^{cold}$  は4月～7月付近での予測精度が他の指標よりも悪くなっている。これは、「寒い」とはほとんどつぶやかれず、十分なデータ数が得られない中で算出されたこの時期の  $S_t^{cold}$  は信頼性にかける指標となっているためであると考えられる。また、平均気温と従来の温熱指標 ( $NET, THI, WCI$ ) を比較してみると、図6-(a)-(c)と同様、平均気温と  $NET$ ,  $THI$  の間に違いはあまり見られないという結果となった。また、 $WCI$  の11月～2月あたりの予測精度が他の指標よりも特に劣っており、寒さの温度感覚に対象を絞ったこの指標は、この期間の「飲料」「乳飲料・清涼飲料」に対しては特に適していないことがわかった。

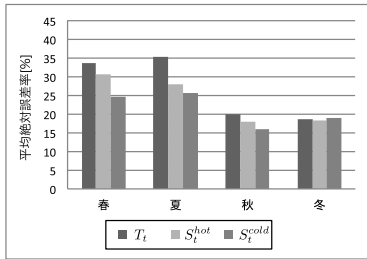
図6より、各指標の月別の誤差傾向については把握できるが、提案指標である  $S_t^{hot}$  と  $S_t^{cold}$  の季節特性を比較するため、四半期ごとの平均絶対誤差率について示した結果を図7に示す(注9)。

一般に、需要予測は夏季と冬季よりも春や秋の季節の変わり目の方が難しいという実情があり、体感気温の導入によって、このような季節の変わり目の時期における需要予測の精度向上が求められている。このことを考慮した上で、図6の結果から読み取れることは以下のとおりである。

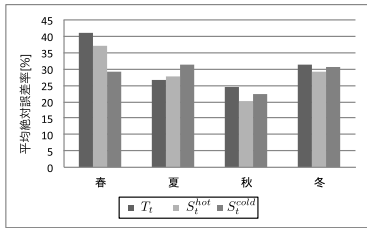
- 寒さに影響を受けやすいと考えられる商品カテゴリー（「菌茸・鍋」、「こんにゃく」、「練り製品」）に対しては、春先から夏にかけての需要予測には  $S_t^{cold}$  の方が精度が高い。秋については商品カテゴリーによって特性が異なり、個別に検討が必要である。

- 暑さに影響を受けやすいと考えられる商品カテゴリー（「飲料」、「乳飲料・清涼飲料」）では、四季を通じて予測精度は高く、秋は  $S_t^{cold}$ 、それ以外の季節

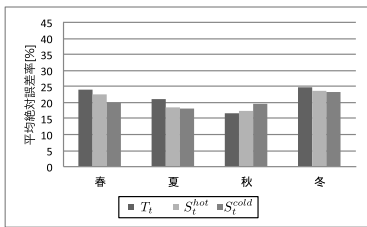
(注9) :  $NET, THI, WCI$  に関しては、図6において平均気温との大きな違いが見られないことから、見やすさのため図7では省略している。



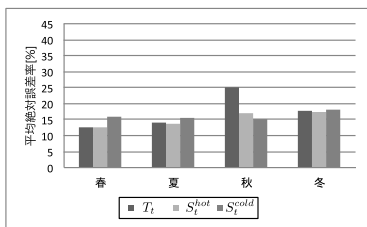
(a) 菌茸・鍋



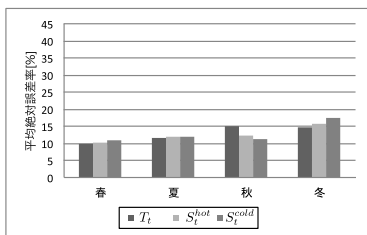
(b) こんにゃく



(c) 練り製品



(d) 飲料



(e) 乳飲料・清涼飲料

図 7 四半期ごとの平均絶対誤差率の比較

Fig. 7 Comparison between mean absolute percentage errors per quarter.

は  $S_t^{hot}$  の方が精度が高い。

すなわち、「菌茸・鍋」、「こんにゃく」、「練り製品」では、春から夏にかけては  $S_t^{hot}$  よりも  $S_t^{cold}$  の精度が優れていることが分かる。一方、「飲料」、「乳飲料・清涼飲料」では、秋以外は  $S_t^{hot}$  の方が精度が優れているが、秋だけは  $S_t^{cold}$  の方が優れているという結果となった。 $S_t^{hot}$  と  $S_t^{cold}$  については、以上のような商品カテゴリーと季節の組み合わせによって需要予測精度の特性の差異が見られ、また夏季や冬季においては平均気温  $T_t$  を用いても同程度の予測精度を示している場合もあるため、これらの手法の使い分けについては慎重に検討すべきである。

また、 $T_t$ +日射との比較に関して、月ごとの精度を確認すると、提案した指標は精度の面で必ずしも優れているということとはできない。しかしながら、例えば「飲料」に関しては、4,5月の春先や9~11月の季節の変わり目で提案指標が優れていることがわかる。このように、提案した体感気温は  $T_t$ +日射と比較しても同程度以上の予測精度を示しており、少なくとも春先や秋口など、時期を限定しての利用価値はあるといえる。また、 $T_t$ +日射のように、気象条件をそのまま説明変数とした回帰モデルでは、多くの説明変数と商品需要の関係性がモデル化されてしまう。そのため、小売業における現場オペレーションレベルの観点からすると、受け入れられ難いという現実もある。様々な気象条件を用いて需要変動を表した予測モデル式が、体感気温のみによる需要予測モデルに置き換われば、例えば、単純に「体感気温が高い日に売れる商品」、「体感気温が低い日に売れる商品」といった極めて明快な商品のカテゴリー化も可能である。この点も、本研究で提案した体感気温を用いるメリットの一つである。次の章では、体感気温を指標として用い、実際に商品の売れ行きとの関係を分析した結果について示す。

## 6. 体感気温と商品の売れ行きの関係の分析

本節では、体感気温と実平均気温の関係に基づいて「暑い日」「通常の日」「寒い日」を定義し、特に「暑い日」「寒い日」に売れ行きが伸びる商品カテゴリーの特定を行う。この結果から、特定された商品カテゴリーに対する細かな在庫管理の一助となる知見が得られることを期待する。

用いるデータは、6. で用いたものと同じ 2013 年 1 月 1 日から 2014 年 12 月 31 日における大手小売チェーンの売上数量データである。体感気温と実平均

気温の関係を基準に、各日に「暑い日」「通常の日」「寒い日」のいずれかのラベル付けを行い、それぞれのラベルに該当する日に特に売れ行きが伸びる商品カテゴリーを特定する。なお、これ以降、第  $t$  日のラベルを  $\ell_t$  と表し、第  $t$  日が「暑い日」のとき  $\ell_t = 1$ , 「通常の日」のとき  $\ell_t = 2$ , 「寒い日」のとき  $\ell_t = 3$  であるとする。

**6.1 「暑い日」「通常の日」「寒い日」の定義**

ここでは、「暑い日」「通常の日」「寒い日」のラベル付けを行う方法について述べる。ここでは、商品需要は実平均気温と比較したときの体感気温の高低差に影響を受けやすいのではないかとという仮説に基づき、この高低差と商品の売上数量との関係に注目する。まず、体感気温  $S_t^{hot}$  と実平均気温の差  $U_t^{hot} = S_t^{hot} - T_t$  を計算し、その一週間移動平均を取る (これを  $U_t^{hot(m)}$  とする)。そして、 $\tilde{U}_t^{hot} = U_t^{hot} - U_t^{hot(m)}$  の値に応じてラベル付けを行う。これは、温度感覚の感じ方は  $U_t^{hot}$  そのものの高低よりも一週間程度の期間内での相対的な  $U_t^{hot}$  の高低に影響されるのではないかとという仮説に基づいている。また、ここまでの手順は  $S_t^{cold}$  についても同様に行い、 $\tilde{U}_t^{cold}$  についても同様に算出する。ラベル付けの定義は以下のとおりである。

$$\ell_t = \begin{cases} 1 \text{ (寒い日)} & \text{if } \tilde{U}_t \leq -a\sigma_{\tilde{U}} \\ 2 \text{ (通常の日)} & \text{if } -a\sigma_{\tilde{U}} < \tilde{U}_t \leq a\sigma_{\tilde{U}} \\ 3 \text{ (暑い日)} & \text{if } \tilde{U}_t > a \end{cases} \quad (13)$$

ただし、 $\sigma_{\tilde{U}}$  は  $\tilde{U}_i$  の標準偏差とし、 $a$  は予備実験等により決定するパラメータとする。また、 $\tilde{U}_i$  は、 $\tilde{U}_t^{hot}$  または  $\tilde{U}_t^{cold}$  を表しており、ここでは  $\tilde{U}_t^{hot}$  と  $\tilde{U}_t^{cold}$  のそれぞれに基づいてラベル付けを行うものとする。

**6.2 体感気温感応度の定義**

ラベルごとの売れ行きの変動を定量化するための指標として、体感気温感応度を定義する。いま、ラベルが  $\ell$  である日の集合を  $\mathcal{T}_\ell := \{t \mid \ell_t = \ell\}$ , 商品カテゴリー  $i$  の第  $t$  日の売上数量を  $n_i^{(t)}$ , 商品カテゴリー  $i$  の期間内の平均売上数量を  $\bar{n}_i$  とする。このとき、 $\bar{n}_{i\ell}$  を

$$\bar{n}_{i\ell} = \frac{1}{|\mathcal{T}_\ell|} \sum_{t \in \mathcal{T}_\ell} n_i^{(t)} \quad (14)$$

と定義する。これはラベルが  $\ell$  である日の商品カテゴリー  $i$  の平均売上数量である。このとき、商品カテゴリー  $i$  の体感気温感応度  $G_i$  を以下のように定義する。

$$G_i = \max_{1 \leq \ell \leq 3} G_i^{(\ell)} \quad (15)$$

$$G_i^{(\ell)} = \frac{\bar{n}_{i\ell} - \bar{n}_i}{\bar{n}_i} \quad (16)$$

$G_i^{(\ell)}$  は、ラベルが  $\ell$  の日の売上数量の平均売上数量に対する増減の割合を表している。また、 $G_i$  の値をとるときのラベル  $\ell$  を  $L_i$  と表記する。

$$L_i = \arg \max_{1 \leq \ell \leq 3} G_i^{(\ell)} \quad (17)$$

実験においては、当該店舗で販売されている全 153 商品のうち売上数量上位 50 件の商品のみについて  $G_i$  を算出するものとする。

**6.3 実験結果**

まず、 $a$  の値を 1, 2, 3 と変化させたときの、各ラベルの付与された日数を表 4 に示す。ここで、 $a$  の値を決める基準としては、「暑い日」「通常の日」「寒い日」の該当日数が年単位で同程度、あるいは「通常の日」が若干多くなるように決定するものとする。前者の基準は、 $\bar{n}_{i\ell}$  の値の評価を公平に行いたいという考えから、後者の基準は、温度感覚について考えた場合、「暑い」や「寒い」と感じる日よりも「どちらとも感じない」日の方が多く、またそのような日の間では需要の変動はそれほど起こらないゆえに先に述べた公平性は失われないのではないかとという仮説からである。表 4 の結果と上記の基準から、 $S_t^{hot}, S_t^{cold}$  のいずれも  $a = 1$  としてラベル付けを行い、体感気温感応度を算出した。

次に、体感気温感応度  $G_i$  が上位 10 件の商品カテゴリー、及びそのときのラベル  $L_i$  を表 5 に示す。表 5 を見ると、 $S_t^{hot}$  と  $S_t^{cold}$  のどちらを用いた場合にも、「菌茸・鍋」、「季節・その他」、「薬物・妻物」といった商品カテゴリーが共通して特に売れ行きが伸びていることがわかる。また、それぞれの商品カテゴリーのラベル  $L_i$  を見ると、寒い日 ( $L_i = 1$ ) には「菌茸・鍋」、「冷凍食品」、「鮮魚」の売上数量が、暑い日 ( $L_i = 3$ ) には、「土物」、「薬物・妻物」、「サラダ」、「季節・その他」の売れ行きが伸びる傾向にあることがわかる。これらの商品

表 4  $a$  の値と各ラベルの付与された日数  
Table 4 The value of  $a$  and corresponding number of days for each label.

$a$ の値	$S_t^{hot}$			$S_t^{cold}$		
	暑い日	通常の日	寒い日	暑い日	通常の日	寒い日
1	187	336	207	202	233	295
2	70	605	55	141	482	107
3	18	701	11	84	636	10

表 5 体感気温感応度  $G_i$  の上位 10 件  
Table 5 Top-10 in  $G_i$ .

$S_t^{hot}$			$S_t^{cold}$		
商品カテゴリー	$G_i$	$L_i$	商品カテゴリー	$G_i$	$L_i$
菌茸・鍋	0.201	1	土物	0.267	3
飲料	0.118	3	サラダ	0.172	3
季節・その他	0.118	3	季節・その他	0.168	3
冷凍食品	0.097	1	菌茸・鍋	0.160	1
葉物・妻物	0.094	3	葉物・妻物	0.134	3
鮮魚	0.094	1	鮮魚	0.089	1
サラダ	0.068	3	冷凍食品	0.074	1
土物	0.066	3	水物・加工野菜	0.049	3
練り製品	0.056	1	とり肉	0.039	1
卵	0.055	1	焼き菓子	0.039	3

カテゴリーは本研究の成果として得られた「気象感応度の高い商品カテゴリー」であり、これらの商品カテゴリーに対して、各日のラベルをもとに  $100G_i\%$  の需要増を見込んだ在庫管理を行うことで、機会損失を減らし、当該商品の売上高を増加させると同時に在庫保管費用の低減が可能となるのではないかと考えられる。

なお、表 5 では、 $S_t^{hot}$  で「菌茸・鍋」が、 $S_t^{cold}$  で「土物」、「サラダ」、「季節・その他」が上位にきている。直感的には、「菌茸・鍋」は寒い日に食べたくなる商品であり、「土物」、「サラダ」、「季節・その他」は暑い日に食べたくなる商品であるため、この点について考察を与えておく。表 5 の結果は、

- $S_t^{hot}$  を用いて得られる「暑くない日」には平均よりも売れ行きが 20.1%ほど伸びる
  - $S_t^{cold}$  を用いて得られる「寒い日」には平均よりも売れ行きが 16.0%ほど伸びる
- ということを示している。ただし、上記の「暑くない日」とは、 $S_t^{hot}$  を用いて式 (13) を適用した結果、 $l_t = 1$  となった日を指し、「寒い日」とは、 $S_t^{cold}$  を用いて論文内の式 (13) を適用した結果、 $l_t = 1$  となった日を指す。この結果は逆に言えば、「菌茸・鍋」の需要への影響は、 $S_t^{hot}$  を用いて定義される「通常の日」、「暑い日」の場合には、需要が大きく落ち込むことを意味している。「菌茸・鍋」の需要への影響は、直感的には「寒い日」かどうかの方が需要への影響が強いとも思われるが、実は「暑くない日」かどうかの方が需要への影響が（若干ではあるが）大きかったと言える。以上の考察は、「土物」や「サラダ」、「季節・その他」に対しても、同様の議論をすることができる。これらの考察は結果から考えられる仮説を含んでいるため、今後、実店舗における実証実験などを通じた検証が必要であり、今後の課題とする。

## 7. 考 察

本章では、本研究で提案した体感気温の定量化のためのモデルに関する考察を行う。本研究における体感気温の定量化において重要視していた点は、気象要素のみから算出が可能であること、体感気温に影響を与える気象要素を特定しそれぞれの影響の度合いを解明すること、得られた体感気温が需要予測に有用であること、の 3 点である。ここでは、この三つの観点から考察を与える。

まず、1 点目について述べる。本研究においては、Tweet 割合の推定値と実測値との差（残差）をそのまま利用して体感気温を算出するのではなく、気象要素による推定値を用いることでこれを実現している。その理由については既に述べているが、このようにすることで得られる別の利点として、Tweet 割合に含まれていると考えられるノイズが気象要素による推定の過程で除かれること、Tweet 割合がおおむね都市部の温度感覚のみを表しているゆえに生じる不整合な部分が緩和されることの 2 点がある。1 点目については、Tweet データの性質上、温度感覚の定量化に用いるのにふさわしくないつぶやき（夏にクーラーの効きすぎた部屋でつぶやかれた「寒い」のようなつぶやきなど）が一定数含まれていると考えられるため、それが気象要素による残差回帰の過程で除かれることが期待できるということである。2 点目については、総 Tweet 数は人口の集中している関東中心部、関西中心部で著しく多いため、Tweet 数から算出される「暑い/寒い」を含む Tweet 割合の値は、同地域で観測された暑さ、寒さの指標であると言わざるを得ない。よって、この Tweet データを用いて、他の地域（A 地区とする）の体感気温を算出しようとした場合、気象データは A 地区のものであるのに対して、温度感覚の指標（＝「暑い/寒い」を含む Tweet 割合）は都心部のもになってしまうという不整合な状況が生じてしまう。しかし、体感気温の定量化モデルの構築の際にのみ Tweet データを用いることで、先に述べた不整合な点が解消され、整合性のとれたモデルを構築することができる。そして、A 地区の体感気温は、得られたモデルに A 地区の気象データを入力するだけで算出できることになる。

2 点目については、各特徴量の重要度が算出可能な Random Forest を用いることで、各特徴量の重要度の値から、体感気温の算出に重要である、つまり体感気温に影響を与える気象要素の特定を可能とした。更

に、それを季節ごとに層別した上で行うことで、季節ごとの影響度合いの違いを見ることができるようにした。この重要度は0から1の範囲で算出されるため、各気象要素が体感気温に正負どちらの方向に影響を与えているかをみることはできないが、重回帰分析により得られた標準偏回帰係数の符号と合わせて統合的に分析するなどの方法により対応が可能であると考えられる。

3点目に対しては、主として体感気温を用いた需要予測の精度が、特に季節の変わり目であり、天気予報の予測が難しい9、10月に特に実平均気温や、これまでに提案されてきた温度感覚に関する指標よりも優れていることを、重回帰分析を用いた需要予測実験において示した。通常気温を用いることでも需要の予測はある程度の精度で可能ではあるものの、季節の変わり目などを対象とした場合、予測が難しい。しかしながら、本提案手法を用いることによって、この時期の予測精度の向上を行うことができた。定量化が難しい体感気温であるが、一方でこの結果は、温度感覚に関わる既存の指標よりも、提案したモデルにより得られる体感気温の方が、需要との関係がより単純な線形関係に近づいたということを示したに留まっている。ゆえに、より厳密に需要予測への利用の有用性を示すには、体感気温から需要予測を行う方法論についても議論していく必要があると言える。これらに関しては今後の課題とする。

## 8. む す び

本研究では、Tweet データと気象データを統合的に用いることで体感気温の定量化を行うモデルを提案した。また、算出した体感気温を用いて気象感応度の高い商品カテゴリーの需要予測を行い、温度感覚に関わる他の指標と比較することにより、提案した体感気温の需要予測への利用の上での有用性を示した。更に、実平均気温と体感気温を用いて、その日が「暑い日」「通常の日」「寒い日」のいずれかであることを定める定義式を提案し、それぞれの日に売れ行きが伸びる商品カテゴリーの特定を行った。その結果、気象感応度が高いと考えられる商品カテゴリーの在庫管理の一助となる知見が得られた。本研究は、大規模 Tweet データと気象データから消費者の感じる体感気温を定量化し、更にそれを考慮して管理を行うべき商品カテゴリーを明らかにすることができたという点で、小売業における気象情報及びその他の大規模データの活用の新たな

観点を示すことができたと言える。

今後の課題としては、温度感覚に関わる単語群の抽出、体感気温の算出に用いる気象要素の検討、需要予測に有用な体感気温の定量化方法の検討、並びに Twitter のデータ以外の方法で消費者の体感気温を算出する方法の検討などが挙げられる。

謝辞 本研究の遂行にあたり、様々な助言を頂いた上智大学の山下 遥 先生、一般財団法人日本気象協会の皆様、早稲田大学創造理工学部 後藤研究室の皆様に深く感謝します。本研究の一部は、科学研究費(26282090, 26560167)の助成を受けたものである。

## 文 献

- [1] A. Missenard, "On thermally equivalent environments," *JIHVE*, vol.27, no.2, pp.231-237, 1959.
- [2] G. Hentschel, "A human biometeorology classification of climate for large and local scales," *Proc. WMO/HMO/UNEP Symposium on Climate and Human Health*, pp.22-26, 1986.
- [3] E.C. Thom, "The discomfort index," *Weatherwise*, vol.12, no.2, pp.57-61, 1959.
- [4] "Wikipedia," <https://ja.wikipedia.org/wiki/%E4%BD%93%E6%84%9F%E6%B8%A9%E5%BA%A6>
- [5] 神山恵三, "体感温度," 気象研究ノート (日本気象学会), vol.12, no.3, pp.214-248, 1961.
- [6] A.P. Gagge, "An effective temperature scale based on a simple model of human physiological regulatory response," *Ashrae Trans.*, vol.77, pp.247-262, 1971.
- [7] A.P. Gagge, A.P. Fobelets, and L. Berglund, "A standard predictive index of human response to the thermal environment," *ASHRAE Trans.*; (United States), vol.92, no.CONF-8606125, pp.709-731, 1986.
- [8] W.E. Woodson, B. Tillman, and P. Tillman, *Human Factors Design Handbook: Information and Guidelines for the Design of Systems, Facilities, Equipment, and Products for Human Use*, McGraw-Hill Education, 1992.
- [9] C.P. Yaglou and D. Minaed, "Control of heat casualties at military training centers.," *Arch. Indust. Health*, vol.16, no.4, pp.302-16, 1957.
- [10] P.O. Fanger, *Thermal Comfort.: Analysis and Applications in Environmental Engineering*, Copenhagen: Danish Technical Press., 1970.
- [11] 木内 豪, "屋外空間における温冷感指標に関する研究," *天気*, vol.48, no.9, pp.661-671, 2001.
- [12] 大橋唯太, "温熱指標 (新用語解説)," *天気*, vol.57, no.1, pp.57-59, 2010.
- [13] 工藤 拓, 山本 薫, 松本裕治, "Conditional random fields を用いた日本語形態素解析," 情報処理学会研究報告自然言語処理 (NL), vol.2004, no.47 (2004-NL-161), pp.89-96, 2004.
- [14] 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.

- [15] F.M. Tseng, H.C. Yu, and G.H. Tzeng, "Applied hybrid grey model to forecast seasonal time series," Technological Forecasting and Social Change, vol.67, no.2, pp.291-302, 2001.
- [16] L. Breiman, "Random forests," Machine Learn., vol.45, no.1, pp.5-32, 2001.  
(平成 29 年 10 月 20 日受付, 30 年 2 月 1 日再受付,  
4 月 2 日早期公開)



馬賀 嵩士

1992 年生。2015 年早稲田大学創造理工学部経営システム工学科卒業。2017 年同大学大学院修士課程修了。在学中、機械学習とその応用に関する研究に従事。



三川 健太 (正員)

1981 年生。2005 年武蔵工業大学環境情報学部環境情報学科卒業。2007 年同大学大学院修士課程修了。2016 年早稲田大学大学院博士後期課程修了。博士(工学)。2013 年早稲田大学助手。2016 年湘南工科大学講師。機械学習とその応用に関する研究に従事。IEEE, 電子情報通信学会, 日本経営工学会など, 各会員。



後藤 正幸 (正員)

1969 年生。1994 年武蔵工業大学大学院修士課程修了。2000 年早稲田大学博士課程修了。博士(工学)。1997 年早稲田大学理工学部助手。2000 年東京大学大学院工学系研究科助手。2002 年武蔵工業大学環境情報学部助教授。2008 年早稲田大創造理工学部経営システム工学科准教授。2011 年同大教授。情報数理応用とデータサイエンスの研究に従事。著書に、「入門パターン認識と機械学習」, コロナ社 (2014), 「ビジネス統計 統計基礎とエクセル分析」, オデッセイコミュニケーションズ (2015) など。IEEE, 電子情報通信学会, 人工知能学会, 日本経営工学会, 経営情報学会など, 各会員。



吉開 朋弘

1986 年生。2010 年京都大学工学部地球工学科卒業。2012 年同大学院大学院工学研究科社会基盤工学専攻修了。2012 年一般財団法人日本気象協会入社。気象レーダを用いた防災分野のデータ解析や気象情報を利用した商品需要予測プロジェクトに従事。