

# カスタマーレビューにおける シズルワードと季節要因の関連性について

佐藤 百恵<sup>†</sup> 小暮 健太<sup>†</sup> 齊藤 史哲<sup>†</sup>

<sup>†</sup> 千葉工業大学先進工学部 〒275-0016 千葉県習志野市津田沼2丁目17-1

E-mail: <sup>†</sup> s19c3055em@s.chibakoudai.jp

**あらまし** シズルワードは食品評価に直結するシズル感を端的に表現できることから、マーケティングの観点からも重要であるといえる。これらを食品に対する評価表現とみなすと、評価対象のニーズと季節性の関連性が高いものがあり、その性質を用例データに基づいて理解することは重要である。例えば、「ほくほく」といった表現は気温が低い秋から冬にかけて「あっさり」などは気温・湿度が高い夏場に好まれる表現といえる。本研究では、気象条件とこれらのシズルワードの関連性を明確にするための低次元マッピングと気象データを対応付けたデータ可視化技術を提案し、実データの解析に適用した。これを用いた解析により、食品評価と気象条件の関連性やシズルワードの利用傾向に関する知見獲得を目指している。

**キーワード** シズルワード, カスタマーレビュー, 季節要因, データ可視化, 不快指数

## 1. はじめに

昨今、コロナ禍での情勢を踏まえて巣籠もり需要が増加し、通販サイトを用いて買い物をする機会が増える中で消費者が商品レビューを投稿する機会も同時に増加している。レビューはその商品の批評する際に利用されるものであり、新規の顧客は商品購入に至るまでのプロセスにおいて意思決定に大きな影響を及ぼしている。

食品のレビューにおいて頻繁に登場する「さっぱり」や「濃厚」等、その食品の特徴を端的に表現し、人々の「美味しそう」や「食べたい」を想起させる言葉を「シズルワード」と呼ぶ[1]。これらは食感や味の印象など、人の感覚的な情報や魅力を伝えるのにあたり重要な表現となっており、広告塔のマーケティング領域で活用されている。

レビューには消費者の商品評価の情報が多分に含まれていることから、レビュー上に登場するシズルワードの傾向を調べることで、消費者の求める食品の特性を知ることができるのではないかと考えた。また、同じ特徴でも消費者の置かれている環境によっては是非が反転することがあると考えられる。そこで月毎のシズルワードの傾向と共に季節的な条件、本研究においては不快指数や湿度を変数に加えて両者の関連性を分析することで、環境によって変化する需要傾向を可視化する。

## 2. 関連研究

シズルワードについての研究は盛んに行われており、加藤ら[2] (2015) の研究では投稿時期によるTwitter上の食品名を含むツイート数の変化やメディア毎に使用されるシズルワードについて調査を行なっている。

延安ら[3] (2016) の研究では、時期とTwitter上に投稿されるシズルワードの関係性について分析を行なっており、時期によって異なる食品とシズルワードの共起関係を調べている。ここでは特に、シズルワードにある3つの領域のうちどの領域が食品と共起するかという視点からも分析を行なっている。

本研究では楽天市場の商品レビューを分析対象とし、登場したシズルワードに対応する食品が明確な中での分析を行う。加えて、先行研究では月毎での比較を行っていたが、もっと具体的な条件としてその月の平均湿度や、不快指数という人間が体感で感じる不快度を表す指数を用いて分析を行う。これらを用い、シズルワードの出現数を分析することで、消費者が食品を消費した際の

環境を仮定することができ、それがどのようにレビューに影響を与えるのかを調査していくことが本研究の目的となる。

## 3. 基礎事項の準備

ここでは、提案法の構成要素について整理する。

### 3.1 不快指数

不快指数とは、Thom[4]によって提唱された温熱指標であり、夏の蒸し暑さを数量的に表したものになる。数値が75以上で不快感を感じ、60以下で寒く感じるとされている。本研究で用いる不快指数を求める数式は以下の通りである。

$$DI = 0.81T + 0.01H (0.99T - 14.3) + 46.3 \quad (1)$$

ここで $T$ は気温(°C)、 $H$ は湿度(%)を表している。

### 3.2 BERT

BERTとはJacob[5]らによって発表された汎用言語表現モデルである。Transformerをモデルに組み込むことで既存のモデルよりも処理速度に優れたモデルを作成できるようになった。

事前学習のみを行なったBERTでは、文脈を考慮しながら入力された文章を768次元のベクトルに変換することが可能となり、これを用いることで文章の類似度を測ることが可能となる。

### 3.3 UMAP

UMAPとは教師なし学習の一種で、McInnesらによって提案された次元削減アルゴリズムである[6]。計算コストが低く処理が早い点がUMAPの特徴の1つとなり、高次元上で距離が近いデータ同士を低次元上でも近く保つことを目的としているが、他のアルゴリズムと比べて似たデータが近くに纏まりやすくなっている。

## 4. 提案手法

### 4.1 基本指針

本研究ではBERTを用いてレビュー文のベクトル化を行い、得られた768次元のベクトルを2次元まで削減、それらを可視化させ分析を行う。

### 4.2 処理手順

以下、本研究での提案手法の処理手順となる。

Step1: レビューデータ内で解析条件に合致したデータを選択し、気象条件データと対応付ける。

Step2: 事前学習済みのBERT モデルを用いて Step1 で選ばれたレビュー文の分散表現を獲得する。

Step3: Step2 で得られた分散表現を 2 次元空間上にマッピングする。

Step4: 気象条件を低次元マップ上に反映し、条件とレビューの対応関係を可視化する。

Step5: クラスタリングによりレビューをグルーピングしてシズルワードや気象条件の傾向を調査する。

## 5. 実験

### 5.1 使用データ

本研究では楽天株式会社より提供された 2019 年 1 月 1 日から同年 12 月 31 日までの楽天市場のカスタマーレビューデータを使用する[7]。対象ジャンルとするのは加藤らの研究により 1 年間で使用されるシズルワードに大きな変化がないとされたパンケーキ、ラーメンを含む洋菓子ジャンルと麺類ジャンルとした。

文献[1]より対象ジャンルに含まれる食品を想起しやすいとされるシズルワードをそれぞれ 6 個ずつ選出し、それらを分析対象シズルワードとして研究を行なった。表 5, 6 は対象となるシズルワードである。

また、季節要因としてはレビューデータと同様 2019 年 1 月から 12 月までの東京の月別平均気温と平均湿度を用いて不快指数を算出した。

表 5 洋菓子ジャンルの対象シズルワード

サクサク	ふわふわ	濃厚
さっぱり	あっさり	コク

表 6 麺類ジャンルの対象シズルワード

新鮮	絶品	濃厚
さっぱり	モチモチ	ツルツル

### 5.2 実験設定

本研究の Step2 でレビューのベクトル化において利用するモデルは東北大学乾研究室が公開する事前学習済みの日本語 BERT を用いている。

獲得した分散表現の次元削減には UMAP を用いる。削減後は 2 次元のベクトルになり、その際の `n_neighbors` の値はそれぞれのジャンルで最適と思われるものを設定し、洋菓子ジャンルでは 9、麺類ジャンルでは 7 とした。

UMAP で次元削減した後、結果に対して k-means++ を用いてクラスタリングを行う。これを結果の解釈の際に補助として使い、傾向を考察するのに使用する。クラスタ数は 20 とし、各クラスタに所属するデータの観察を行う。

### 5.3 結果と考察

#### 5.3.1 シズルワードを含むレビュー投稿数の比較

図 1 は洋菓子を対象とした「さっぱり」出現数と不快指数の月別グラフで、夏の不快指数が高い時期に出現数が増加している。他にも図 2 では不快指数が低い時期に出現数が増加しており、図 3 では通年変化が見られず同じ商品特性でも対象となる食品によって求められ方に違いが出るのが読み取れる。

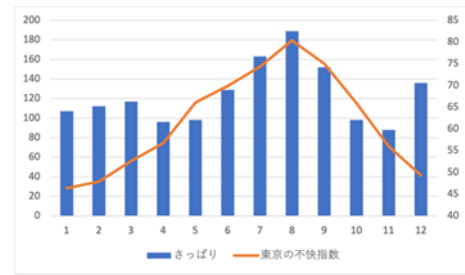


図 1 「さっぱり」(洋菓子) 出現頻度と不快指数

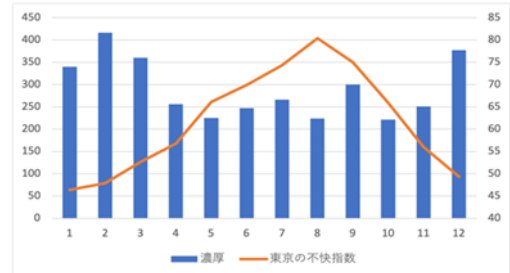


図 2 「濃厚」(洋菓子) 出現頻度と不快指数

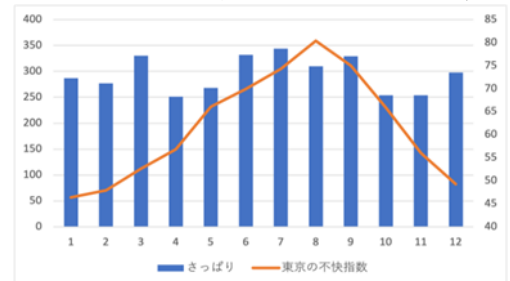


図 3 「さっぱり」(麺類) 出現頻度と不快指数

#### 5.3.2 BERT での低次元化

続けて、BERT を用いて得られた文章ベクトルから 散布図の作成を行なった。図 4 は洋菓子ジャンルにおけるレビューの低次元マッピング結果をクラスタリングした中で、1 つのクラスタを黄色で描画したものとなる。図 5 をみると、不快指数 60 以下のデータが 64.9%と、人々が寒いと感じる時期に投稿されたレビューが多く含まれるクラスタとなっている。しかし、シズルワードに偏りは見られず、寒い時期のレビューは内容が似通ったものになる可能性がある。

図 6 は麺類ジャンルにおけるレビューの低次元マッピングであり、黄色で示されたクラスタ内では梅雨から夏の時期にかけてのレビューデータが多くなっており、図 7 では、不快指数 70 以上が 30.8%となった。それとは別に梅雨入りが始める 6 月のデータが 16.1%と一番高い割合になっている。シズルワードは「絶品」と「ツルツル」がそれぞれ 20%以上を占め、洋菓子ジャンルとは異なり時期とシズルワード両方がレビューの内容に影響を及ぼしていることが確認できた。

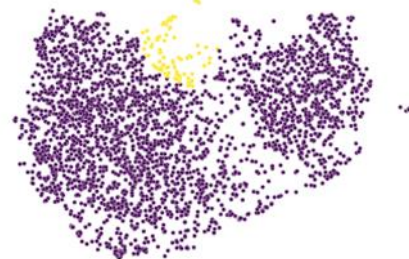


図 4 洋菓子ジャンルの冬のクラスタ

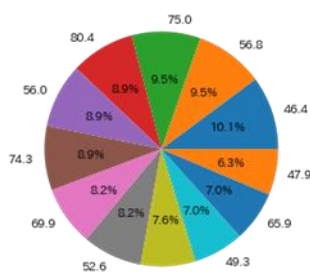


図5 図4でのクラス内不快指数

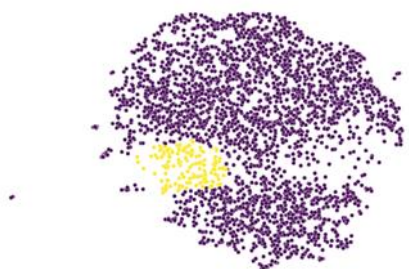


図6 麺類ジャンルの夏のクラス

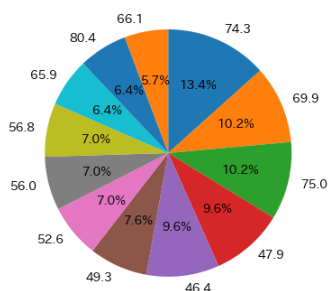


図7 図6でのクラス内不快指数

## 6. まとめと今後の課題

シズルワード毎のレビュー数の比較においては、季節によって変化が出るワード、出ないワードの確認ができた。レビューの低次元マッピングを行うと、一部のクラスにおいて気象情報の影響を受けたクラスを確認することができ、投稿者の置かれている環境とレビューの内容に関係性があることがわかった。しかし、他のクラスにおいてはこうした影響が見られない部分もあり、今後の課題としては他の季節要因（気温や湿度、共起単語など）との関係性を探る必要がある。

## 謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社様から提供された「楽天データセット」を利用しました。この場をお借りして厚く御礼申し上げます。また本研究は科学研究費（基盤 C）19K04887 より支援いただきました。

## 参考文献

- [1] 大橋正房, 光岡祐子, 須藤正彦, 北村奈津子, 米山真知子, 猿渡一兵, BMF ことばラボ” sizzle word シズルワードの現在 2015「おいしいを感じる言葉」調査報告”, BMFT 出版部 (2015).
- [2] 加藤大介, 宮部真衣, 荒牧英治, 灘本明代, ” インターネット上のメディア毎の「おいしさ」表現比較分析”, DEIM Forum 2015 F5-5, (2015).

- [3] 延安幸広, 加藤大介, 宮部真衣, 荒牧英治, 灘本明代, “投稿時期に着目した Twitter 上のシズルワードと食品の関連分析”, 情報処理学会代 78 回全国大会, pp.1-537-1-538, (2016).
- [4] Thom Earl Crabill, “The Discomfort Index”, Weatherwise, Vol.12, pp57-61, (1959).
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, NAACL-HLT 2019, pp.4171-4186, (2019).
- [6] Leland McInnes, John Healy, James Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension education”, arXiv, 1802.03426 (2018).
- [7] 楽天グループ株式会社 (2014): 楽天データセット, 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.0>.