

産直 EC サイト「食べチョク」における季節キャンペーンの効果改善：文字ベクトルモデルを活用して

菊田祐太郎^{†1} 佐藤瑠一^{†2} 赤木徳顕^{†1}

キーワード：購買予測モデル、文字ベクトル、カスタマージャーニー、AUC

1. はじめに

ユーザー数 95 万人を抱え、4 年連続で日本国内の産直通サイトでの業界トップとなっている「食べチョク」を運営するビビッドガーデン社（以下ビビッド社）と帝京大学の間において同社の優良顧客分析、特に季節商品キャンペーンにおいて、同社の優良顧客はどのような購買行動を示すのか？キャンペーンに反応するのはどのような顧客なのか？に関して調査を進めることに合意し、共同研究を行っている。ビビッド社の関心としては、季節ごとに行う果物などの季節限定商品を新規顧客に購入して貰う中で、同サイトの優良顧客は、商品検索において特定品種の名前で検索するなど「ワード」に拘った利用内容が確認されている。

そこで本研究においては「文字ベクトル」に着目し、キャンペーン対象商品にたどり着く前の直近 10 商品の商品名を文字ベクトル化し、その数値を彼らのキャンペーン対象商品に対する関心度として捉え、今後それらの商品の購買予測モデルの構築を行った。購買予測モデルの構築の流れとして、まずはキャンペーン対象商品にたどり着く前の直近 10 商品の商品名のデータ計 8728 人分を文字ベクトル化する。そして、作成した文字ベクトルと特定の単語ベクトル間がなす角のベクトル値から類似度を算出する。その後、類似度から購買予測モデルを作成し、AUC によって評価を行う。

2. 先行研究

我々が注目した先行研究は以下の通りである。単語・文章ベクトルに関する先行研究として、Yamagiwa ら[1]は、データを独立な成分に分解する独立成分分析（ICA）を利用することで、単語や画像の埋め込みに存在する「解釈可能なトゲ」を見つけることができ、ICA で明らかになる埋め込みの普遍的な形状について明らかにした。

食品への知覚や表現の言語表現に関する先行研究として、加藤ら[2]は、食品と美味しさを表す単語のメディアごとの使用傾向を明らかにした。陳ら[3]は、りんごの味覚成分や物理特性から美味しさの客観的評価の傾向を明らかにした。

3. フレームワーク

近年注目を集める AI 分野の技術のなかに「自然言語処理」があり、これを活用することで単語や文章の意味的な強さを、定量化してベクトルとして表現することができる。こうして算出された単語ベクトルや文章ベクトルが複数ある場合、ベクトル同士がなす角のコサイン値やユークリッド距離を用いることで、単語や文章間の類似度を計算することができる。この技術は検索エンジンや AI チャットボットなどに組み込まれ、ビジネスやマーケティングの現場に活用されている。

本研究では単語・文章ベクトル間の類似度の計算によって、企業が実施するキャンペーンが顧客の購買行動にどのような影響を及ぼすのかを客観的に明らかにすることを目的とした。研究方法として、キャンペーンページへ行き着くまでの 10 個の検索経歴を文章ベクトル化し、作成した文章ベクトルと特定の単語ベクトルとの類似度を求め、比較した。これにより、どのような単語に関心を持つ人が、キャンペーンページをクリックするのかを探求した。

先行研究などの分析を経て、我々は顧客がキャンペーンに辿り着くまでのカスタマージャーニーを分析するべく、キャンペーンページに辿り着く直前までに辿った商品名データを 10 個用意し、それらを合計 8728 人分のデータを用いて分析を行った。それらの商品名を文字からベクトルに変換する為に Word Embeddings の手法を利用してベクトル化を行った。ベクトル化を行う際に対象のキーワードを「リンゴ」とし、リンゴと 1 セット 10 個の商品名から構成される文書データの類似度を算出した。算出された類似度を用いて、りんごを購入するかしないかのモデルの構築を最終的に行った。

Features	Label
0 【天然本マグロ】漁師激賞！なめろう&漬け(100g×2)♪栗 ポロタン♪【限定2セ	0
1 「雷国から季節限定のお届け物、越冬野菜セット。あま〜い白菜・大根・キャベツ」兵庫県香	0
2 【夏ギフト】2024年8月下旬 TBS ラヴィット！紹介1万本売れて話題沸騰中『富士信玄とうもろ	0
....
8725 リッチな黄金じゃがいも！インカのめざめ 5kg 【特別栽培】【小玉】酸っぱいだけじゃない！	1
8726 信州小布施 シャインマスカット 2房 (約1キロ) 🍓【家庭用】青森産とらずふじ小玉多め9〜	1
8727 食べきりサイズ♪干し芋紅はるか半生タイプ50g×8袋入【冬ギフト】太陽と海のまどんな	1
8728 rows x 2 columns	

図 1 文字ベクトル計算に用いる文章データ

†1 帝京大学

†2 岩手大学

4. 現時点での分析結果

総データ数 8728 人分を文字ベクトル化した類似度の平均値は 0.41 となり、最大値 0.596、最小値が 0.128 となった。また、この類似度を基に対象記事に訪問後にリンゴを購入したかどうかをダミー変数として表したデータ (1: 購入, 0: 未購入) を用いて購買予測モデルを構築した結果、AUC 値は 0.61 となった。

5. 考察

AUC の数値としては目覚ましい結果を未だ得られてはいないが、ビビッド社が現状使っているモデルも 0.65 程度であることから、これから述べる考察次第でより実務的に有益なインサイトが得られると研究を進めている。技術的な課題としては、りんごとりんごの品種が文字ベクトルモデル上で結びつかなかった点が問題点として挙げられる。この結果から、農産物商品を定量的なアプローチによって予測する難しさが浮き彫りとなった。その他にも、食べチョクの利用者の特殊性から、通常の EC サイトやスーパーマーケットのような数量データを用いた購買予測が非常に難しいことも挙げられる。今後は利用者に対して行うアンケート調査ベースの定性データを取り扱った調査を行っていききたい。

参考文献

- [1] Brynjolfsson, E., *The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence*, Dædalus, (2022)
- [2] 鶴光太郎. “新たなテクノロジーは働き方をいかに変えるか,” RIETI PDP 19-P-023 (2019).
- [3] Cognizant, “21 Jobs of the Future “. <https://www.cognizant.com/us/en/whitepapers/documents/21-jobs-of-the-future-a-guide-to-getting-and-staying-employed-over-the-next-10-years-codex3049.pdf>, (2022 年 12 月 15 日アクセス).
- [4] Kotler, P., Kartajaya, H., Setiawan, I., *Marketing 5.0 : Technology for Humanity*, Wiley, (2021)
- [5] 研究開発戦略センター. “人工知能研究の新潮流” CRDS-FY2021-RR-01, (2021)