

# 購買アイテムを特定する分類器パラメータを用いた商品分析に関する一考察

山極 綾子<sup>†</sup> 後藤 正幸<sup>‡</sup>

早稲田大学創造理工学研究科<sup>†</sup> 早稲田大学<sup>‡</sup>

## 1. はじめに

近年 EC サイトを通じた購買行動によって蓄積される購買履歴データの活用は一般的となりつつあり、これらのデータに基づく商品分析手法として、Item2Vec[1]などの埋め込み手法が注目を集めている。これらの手法は、顧客と商品の共起関係に着目し、同じ顧客に購入された商品は類似していると見なすことで、それら商品が近くに埋め込まれるような特徴空間を生成するものである。さらに EC サイト上で購入される商品について、従来はリアル店舗での購入が主であった贈答品や耐久消費財などの商品の購買についてもオンライン上で行われるようになってきている。しかしこれらの商品は顧客の購買頻度が低く、さらに購入目的が商品選択に与える影響が非常に大きいという特徴を持つ。例えば贈答用の商品について、「お見舞いの品として菊の花は不適切」といったように、顧客の嗜好以上にそれらの商品がその目的に照らして適切であるかが考慮される場面が多い。

上記のように購買頻度が低く、かつ同一顧客の連続した購買であってもその個人の嗜好が反映されていると見なすことが難しいデータを対象として、分類器の係数を利用した手法[2]が筆者らによって提案され、実データに対して一定の成果が得られている。具体的には、各商品を購買するか否かを特定する二値分類器を学習し、推定されたパラメータを用いた商品分析モデルを提案している。本稿では人工データを用いて筆者らの手法の特性を明らかにし、更なる応用の可能性を探るとともに、その有効性を示す。

## 2. 準備

### 2.1 推定パラメータによる商品分析モデル[2]

購買頻度が低く、かつ購買の度に顧客の嗜好が変化する商品を対象として筆者らが提案した商品分析モデルについて、その概要を3つの段階に分けて以下に示す。

まず、二値分類器を学習する際のデータセットを作成する。上記手法ではある商品の購買有無を識別する分類器を学習するため、対象の商品の購買データを正例として扱う。ここでその他の商品の購入データを全て負例として扱った場合、データ数が極端に偏ってしまうため、負例のサンプリングを行う必要がある。ただし、その商品の購買有無が購入目的に依存している場合、得られる識別境界が影響の強い購入目的に強く依存してしまうため、顧客嗜好の表現が難しくなることに注意が必要である。そのため、顧客の嗜好を反映できるような負例を選択する必要があるといえる。

次に、各購買履歴における顧客属性や購入目的を説明変数、ある商品を購入したか否かを目的変数とした二値分類器を学習する。文献[2]では、分類器として Factorization Machine(以下、FM)を用いた。これは、FM を用いて推定される係数が目的変数である各商品と、説明変数である顧客属性などの関係性を示しており、それらの解釈が容易であるためである。なお、同様にパラメータの解釈が可能な手法として重回帰モデルも挙げられるが、本研究の対象事例においては、FM を用いて交互作用を考慮することの有効性が示されている。

最後に、得られた係数を用いて商品の類似度分析を行う。文献[2]では各商品においてその商品が購買されたデータを正例として学習された二値分類器の係数を商品の特徴ベクトルと見なし、それらが類似した商品を類似商品として分析している。実際に生花 EC サイト上のデータを対象として分析を行った結果を図1に示す。



図1. ある商品を対象とした類似商品分析

なお、赤枠は同じ花材を用いていることを示し、右下の単語はその購入用途を表す。図1より、同様の花材を用いて作成されており、かつ他の購入目的用の商品の類似度が高くなっており、手法の有効性を示すことができたといえる。

Evaluation of analysis model for products with coefficients of binary classifiers to identify the purchase

<sup>†</sup> Ayako Yamagiwa, School of Creative Science and Engineering, Waseda University

<sup>‡</sup> Masavuki Goto, Waseda University

## 2.2 人工データの生成

本稿では、2.1 節に述べた手法の特性を検証するために、人工データを生成し実験を行った。本節では実際に生成したデータの詳細について説明する。i番目の商品 $p_i$ について、それぞれ直接効果を表す特徴ベクトル $\alpha$ と交互作用を表す特徴行列 $\beta$ を定義し、商品はその属するクラスごとに類似した $\alpha$ と $\beta$ を持つものとする。本実験では特徴量の次元数 $D = 15$ 、カテゴリ数は3、各クラスに属する商品は20とした。各商品の購買確率に対する各特徴量の影響を示す直接効果の値を図2に示す。なお、商品番号0-20はクラス1、21-40はクラス2、40-60はクラス3に属しているものとする。

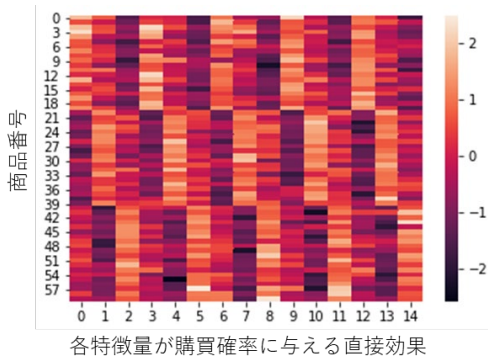


図2. 生成された係数 (直接効果)

次に、生成した特徴ベクトル $\alpha$ 、 $\beta$ を用いて購買履歴データを生成する。各購買における特徴量 $\mathbf{x} = \{x_1, \dots, x_D\}^T$ に基づき、各商品の購買確率 $P_i(\mathbf{x}, \alpha^i, \beta^i)$ を以下の式(1)、(2)で求める。

$$P_i(\mathbf{x}, \alpha^i, \beta^i) = 1 / (1 + e^{-f_i(\mathbf{x}, \alpha^i, \beta^i)}) \quad (1)$$

$$f_i(\mathbf{x}, \alpha^i, \beta^i) = \sum_{d=1}^D \alpha_d^i x_d + \sum_{d=1}^D \sum_{d'=d+1}^D \beta_{d,d'}^i x_d x_{d'} \quad (2)$$

購買履歴データ生成では $P_i(\mathbf{x}, \alpha^i, \beta^i)$ に従い購買商品を確率的に選択し、5000件を作成した。

## 3 実験

### 3.1 概要

本稿では、以下の2つの観点から検証を行う。

- i. 二値分類器で推定された係数の妥当性
- ii. 係数を活用した類似度分析の妥当性

具体的には、2章で述べた人工データを活用し、生成した購買履歴データを用いて得られた係数が、真の値にどの程度近いかどうかを分析する。

### 3.2 結果

初めに、FMを用いて推定された係数 $\alpha$ を図3に示す。なお、二値分類器を学習する際の負例として正例と同じ数だけ、正例以外のすべての購買データからランダムサンプリングを行った。

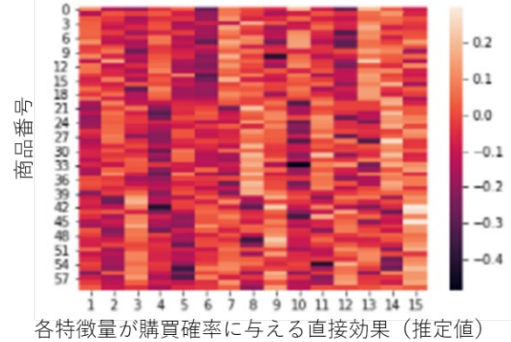


図3. 推定された係数 (直接効果)

図2, 3より、推定された値の絶対値は異なるが、傾向はある程度捉えられており、二値分類器により各特徴量が商品の購買に与える影響を推定可能であると考えられる。さらに推定された係数に対してクラスターリングを行い、真のクラスターの推定が可能かどうかについて検証を行った。その結果、全てのクラスターを正しく推定することが可能であった。従って、係数を用いて類似度分析を行うことは妥当であったと考えられる。

## 4 考察

2章で述べた人工データにおいては、ユーザ情報が含まれておらず、そのためユーザと商品の共起に基づく埋め込み手法の適用は難しい。一方で、個々の購買に着目する筆者らの手法では、真の係数がある程度推定することが可能となっており、その有効性を示していると判断できる。

## 5 まとめと今後の課題

本稿では、筆者らの提案した二値分類器を用いた商品分析手法について、人工データからその特性を検証した。その結果、任意の条件下においては適切に係数を推定することが可能であることが示された。今後の課題としては、真の特徴ベクトルとその推定精度の関係性についての分析と検証が挙げられる。

### 謝辞

本研究の一部は、日本学術振興会 (JSPS) 科学研究費助成事業 No. 21H04600 の助成を受けたものです。

### 参考文献

- [1] Barkan, Oren, and Noam Koenigstein. "Item2vec: neural item embedding for collaborative filtering." *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2016.
- [2] 山極綾子, 雲居玄道, & 後藤正幸, "生花 EC サイトの購買履歴に基づく商品特性分析モデル". 電子情報通信学会論文誌 Vol. J105-D, No. 05, 掲載決定, 2022.