

データサイエンス

key word

- 統計学
- 統計解析
- 非負値行列因子分解
- テキスト解析



佐藤 健一
Kenichi Satoh

データサイエンス学部
教授

【プロフィール】

- ・広島大学大学院理学研究科（博士課程前期）数学専攻 修了
- ・広島大学原爆放射線医科学研究所 計量生物研究分野 准教授
- ・滋賀大学データサイエンス学部 教授 2021年4月～継続中

【主な社会的活動】

日本統計学会, 応用統計学会

【主な受賞歴】

- ・2010年度, 2015年度および2022年度応用統計学会 学会賞(優秀論文賞)受賞

【連絡先】

kenichi-satoh@biwako.shiga-u.ac.jp



ksatohds.github.io

【代表的な研究テーマ】

□ 統計学

課題解決に役立つシーズの説明

買い物データから「好み」を見つけてみよう

— AIを支える技術「NMF(非負値行列因子分解)」のしくみ —

滋賀大学 データサイエンス学部 佐藤 健一



Amazonや楽天の「あなたへのおすすめ」は、どうやって決まるの？ たくさんの人の買い物データから、AIが「好みのパターン」を見つけているのです。ここでは6人のお客さんと6つの商品の小さなデータを使って、AIを支える技術「NMF(非負値行列因子分解)」のしくみを覗いてみましょう。

① まずは「買い物の表」を眺める

お客さん	ケーキ	チョコ	クッキー	サラダ	ヨーグルト	フルーツ
明子	1	1	1	0	0	0
健太	0	0	0	1	1	1
花子	1	1	0	0	0	0
大輔	0	0	0	1	0	1
美咲	1	0	1	0	1	0
太郎	0	1	0	1	1	1

「買った = 1, 買ってない = 0」だけで記録した たった36個の数字の小さな表です。

この表をよく見ると、ある「気づき」がありませんか？

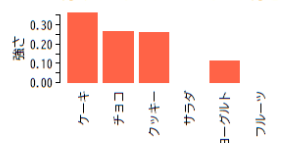
・明子さんは ケーキ・チョコ・クッキー → 甘いもの好き？

・健太さんは サラダ・ヨーグルト・フルーツ → 健康志向？

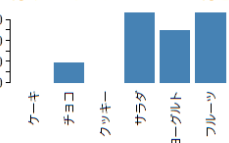
似ているお客さんの「組」が見えてきます。この「似ている」をAIに見つけてもらいましょう。

② NMFが見つけた「2つの好み」

好みのパターン1: スイーツ好き



好みのパターン2: ヘルシー好き



NMFに「2つの好みを見つけて」とお願いするだけで、**スイーツ好き**と**ヘルシー好き**の2パターンが自動で見つかりました。

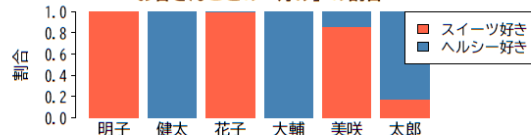
■ パターン1

ケーキ・チョコ・クッキーが高い

■ パターン2

サラダ・ヨーグルト・フルーツが高い

お客さんごとの「好み」の割合



お客さんごとに「好み」を割合で表せます。

・明子さん・花子さん → ほぼ100%スイーツ好き

・健太さん・大輔さん → ほぼ100%ヘルシー好き

・美咲さん・太郎さん → 両方が混ざっている

「好きの種類」と「混ざり具合」を割合で表せるのがNMFの特徴です。

③ NMFが予測する「おすすめ商品」

お客さん	ケーキ	チョコ	クッキー	サラダ	ヨーグルト	フルーツ
明子	-	-	-	0	0.3	0
健太	0	0.3	0	-	-	-
花子	-	-	0.6	0	0.2	0
大輔	0	0.2	0	-	0.6	-
美咲	-	0.7	-	0.1	-	0.1
太郎	0.3	-	0.2	-	-	-

NMFは、まだ買っていない商品ごとに「おすすめ度」を予測できます。

・「-」は もう買った商品 → 無視

・「0」は 買いそうにない → 無視

・数字が大きいほど おすすめ度が高い

NMFのおすすめ

花子さん → クッキー

大輔さん → ヨーグルト

美咲さん → チョコ

NMFが教えてくれること

たった36個の数字から、「好み」「混ざり具合」「おすすめ」が見えました。

これはAIの世界では「パターンの発見」と呼ばれ、画像認識・音声認識・生成AIにも使われている考え方です。

企業・自治体へのメッセージ

- ・アカデミックな共同研究の経験が豊富です。
- ・統計ソフト「R」を使って色々な統計解析手法を使ったり、説明したりできます。