O-022

正則化項付き非負値行列因子分解を用いた POS データにおける商品の関連性分析
Product Relevance Analysis from POS Data Using Non Negative Matrix Factorization with Regularization

濵屋聡太† 南野友香† 森山卓† 細江美欧† 桑野将司†
Sota Hamaya Yuka Minamino Taku Moriyama Mio Hosoe Masashi Kuwano

1. はじめに

さまざまな企業や組織が，革新的なサービスの開発や業務効率の向上を目的に，ビッグデータの活用を始めている．企業や店舗におけるビッグデータ活用の例として，POS データに基づく販売施策がある．POS データを分析することで，売れ筋の商品や，客層ごとに好まれる商品の特徴などを把握できる．しかし，数多くある商品の売上傾向を 1 つ 1 つ分析し，その特徴を把握することは，膨大な計算時間が必要であり実用的ではない．さらに，商品間の売れ行きには，ある商品が売れると別の商品も売れる，あるいは，ある商品が売れるとほかの商品が売れなくなるというような関連性が存在するため，商品単体で分析を行うと，商品間に内在する関連性が無視されることになり，誤った解釈を招く危険性がある．

複雑に関係し合う複数事象間のデータの関連性を明らかにする統計手法の 1 つとして非負値行列因子分解がある．これは，多変量データに潜む共通因子を探り出すための手法であり，多数ある変数を少数の変数，すなわち類似したパターンに分類し，抽出する．近年，さまざまな分野で適用がすすんでいる．原田ら [1]は，携帯電話の位置情報データであるモバイル空間統計データに対して正則化付き非負値行列因子分解を適用した．これにより，年代別および全体における行動パターンの特徴を抽出することに成功した．

本研究では，ノイズ項を取り除き，特徴を容易に解釈するために正則化法を導入した非負値行列因子分解を，各商品の売上データに適用することで，商品分類と売上傾向を同時に抽出することを目的とする．これにより、ある小売事業者の店舗リニューアルや緊急事態宣言による売上傾向の変化把握を試みる．

2. 使用データの概要と分析対象商品の選択

本研究では，ある小売事業者の POS データを用いる．データ取得期間は 2018 年 4 月 1 日から 2020 年 9 月 30 日の 914 日間である．継続的に購入された商品間の関連性を明らかにするため，POS データから毎月 4 個以上購入された商品を分析対象とした．ただし，緊急事態宣言により店休日となった 2020 年 4 月，5 月に関しては，両月の営業日の合計が 13 日となったため，この期間内に 2 個以上購入された商品を継続的に購入された商品と定義した．その結果，分析対象とする商品は 92 種となった．

3. 商品間の関連性に着目した売上傾向分析

3.1 正則化付き非負値行列因子分解の概要

本研究では，商品間の関連性を考慮した売上傾向を把握することを目的に，正則化項付き非負値行列因子分解を用いる．非負値行列因子分解とは，非負値のみを要素にもつ行列データを分解し，情報量を縮約するための分析手法である．非負値行列因子分解は，非負値のみで構成されるI行J列の行列Xを，I行K列の基底行列UとK行J列の表現行列Vの積により近似する．ここで，行列Xを分解するための因子数はKで表される．本研究では，取り扱うデータの構造に合わせ，基底行列Uを時系列パターン，Vを商品パターンと呼ぶ．非負値行列因子分解は，式(1)のように目的関数を行列Xと行列積UVの差とした最小化問題を解くことで行列を分解する．

$$\min_{U\geq 0,V\geq 0} \|X-UV\|_{\mathrm{F}} \tag{1}$$

ここで，$\| \ \|_F$はフロベニウスノルムを表す．

さらに，本研究では，非負値行列因子分解に正則化法を用いる．正則化法とは，ラッソ回帰やリッジ回帰で回帰係数を推定する際に，回帰モデルに罰則項を付与し，モデルの複雑さを低減させるための手法である [2]．また，正則化法は機械学習における過学習を防ぐための代表的な手法としても知られている．本研究では，膨大な商品に関する売上データから抽出されたパターンの特徴を容易に解釈し，売上傾向の顕著な特徴を把握するために，正則化法により時系列パターンおよび商品パターンから不要なノイズ項を取り除く．Kaya et al.（2018）[3]を参考に，商品パターンに L_1 ノルム罰則化項を付与した正則化項付き非負値行列因子分解を式(2)に示す．

$$\min_{U\geq 0,V\geq 0} \|X-UV\|_F+\lambda\|V\|_1 \tag{2}$$

ここで，λはハイパーパラメータ，$\| \ \|_1$は L_1 ノルムを表す．

3.2 因子数，ハイパーパラメータの決定

本研究では，式(3)で表すXとUVの差である 2 乗誤差Eに基づき，因子数Kを決定する．

$$E=\|X-UV\|_F \tag{3}$$

因子数Kは，解釈の容易さと圧縮精度の観点から$K=3$から$K=11$の範囲で推定した．本研究では，因子数Kとハイパーパラメータλの決定について，誤差，MAPE，商品数，各因子が説明する商品数の 4 つの指標を用いた．表 1 に因子数Kにおける誤差Eの最小値と，ハイパーパラメータλ，MAPE，商品数を示す．ただし，MAPE は店休日などにより売上がない日を除いて計算した．商品数については，各因子において算出した時系列パターンの平均と商品パターンの積が 1 以上となったものを，その因子において影響力の強い商品として抽出した．複数の因子で抽出された商品

†鳥取大学　Tottori University

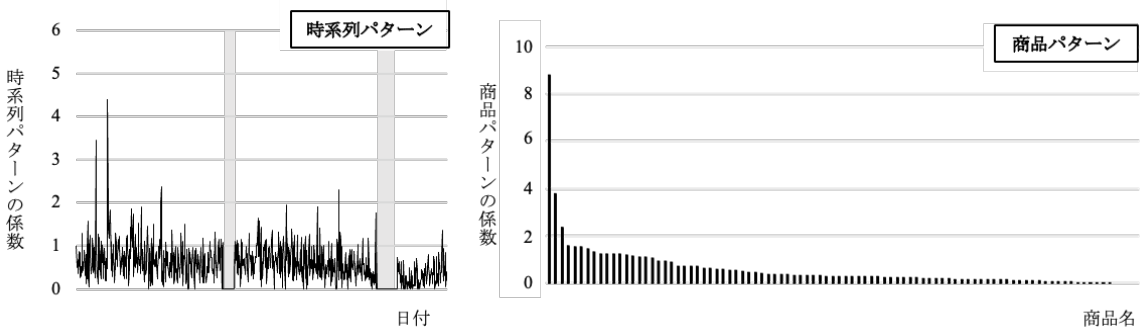

図 1　第 1 因子の時系列パターンと商品パターン

表 1　指標（誤差E，MAPE(%)，商品数）

K	λ	誤差E	MAPE(%)	商品数
3	10^7	489,251	7.72	29
4	10^9	452,811	6.86	27
5	10^6	417,940	6.82	27
6	10^8	385,527	6.72	24
7	10^6	360,209	6.32	26

は，商品数を 1 として計上した．因子が抽出する商品数は，多すぎると売上傾向の解釈が困難になり，少なすぎると得られる情報が僅かになるため，適当な因子数を適切に決定する必要がある．

表 1 より，因子数Kを増加させると，誤差Eの最小値，MAPE が減少し，圧縮精度が向上する．因子数$K = 6$のときは，各因子に配分される商品数が大きく変化したため，因子数Kは 5 以下が適切と考えられる．したがって，分析結果の解釈の容易さを考慮し，本研究では因子数$K = 5$，ハイパーパラメータ$\lambda = 10^6$の場合の考察を行う．

4. 分析結果と考察

図 1 は，抽出された 5 因子のうち，一例として，第 1 因子の時系列パターンと商品パターンを示したものである．時系列パターンにおいて，灰色で示した 2 つの期間は順に，リニューアルオープンのための店休日，緊急事態宣言の影響による店休日をそれぞれ示す．商品パターンは，売上の多い商品順に並べ替えたものを示す．第 1 因子は商品数が 17 であり，飲料，特産品，菓子類が最も多く抽出された．売上は減少傾向であるが，緊急事態宣言後の売上は回復傾向にある．また，突発的に売上が増加する日がある．第 2 因子は商品数が 5 であり，ちくわを中心に抽出された．大きな変動が少なく，比較的安定した売上推移を示すパターンである．緊急事態宣言以前は売上が減少傾向にあったが，宣言後は徐々に回復している．第 3 因子は商品数が 5 であり，商品群にはばらつきが見られた．売上は安定した傾向を示した．緊急事態宣言後，売上が大きく増加する日が見られる．第 4 因子は商品数が 5 であり，ちくわ類，菓子類が最も多く抽出された．売上は減少傾向を示し，突発的に売上が増加する日が見られた．第 5 因子は商品数が 7 となり，野菜類が抽出された．売上は増加傾向を示した．

さらに，時系列パターンを比較するために，データ期間を 3 分割した．第 1 期間をリニューアルオープン前である 2018 年 4 月 1 日から 2019 年 4 月 25 日，第 2 期間をリニューアルオープン後から緊急事態宣言前の 2019 年 4 月 26 日から 2020 年 5 月 31 日，第 3 期間を緊急事態宣言後の 2020 年 6 月 1 日から 2020 年 9 月 31 日と設定した．図 2 に示すように，全体の売上は減少傾向であることがわかり，第 5 因子でのみ 3 期間を通して，売上が増加傾向であることが明らかとなった．

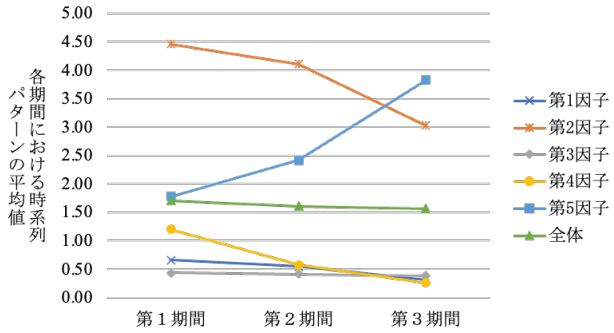

図 2　期間別売上推移

5. おわりに

本研究では，POS データにおける特徴的な時系列パターンおよび商品パターンの抽出を目的に，ある小売事業の POS データに正則化項付き非負値行列因子分解を適用した．提案手法により，より顕著な特徴を示す時系列パターンおよび商品パターンの抽出を試みた点に新規性がある．分析の結果，92 商品の売上傾向を 5 種類のパターンで説明できることが明らかとなり，情報量の圧縮に成功した．また，個別に商品の売上傾向を見た際に把握が困難な商品間の関連性を確認できた．今後の展望としては，特売日の設定や，商品の関連性を考慮した陳列方法などの新たな販売施策決定への重要な知見となりうることが期待される．

謝辞

本研究はある小売事業者よりデータの提供を賜り実施した．ここに記して謝意を表する．

参考文献

[1] 原田魁成，山口裕通，寒河江雅彦，“スパース非負値行列因子分解を用いた COVID_19 流行期の県間旅行行動の変容分析”，土木学会論文集 D3（土木計画学），Vol. 77，No. 2，pp. 160-173，2021.

[2] P.O. Hoyer, “Non-negative sparse cording neural networks for signal Processing XII”, *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, pp. 557–565, 2002.

[3] O. Kaya, R. Kannan, and G. Ballard, “Partitioning and Communication strategies for Sparse Non-negative Matrix Factorization”, *Proceedings of the 47th International Conference on Parallel Processing (ICPP2018)*, pp. 1-10，2018.