

外部条件を考慮した小売店における商品別売り上げの要因分析モデル A Sales Factor Analytical Model for Products in Retail Stores Considering External Conditions

平野洋介* 御供信薫* 楊添翔* 後藤正幸* 吉開朋弘†
Yosuke Hirano Nobushige Mitomo Tianxiang Yang Masayuki Goto Tomohiro Yoshikai

1 はじめに

多種多様な商品(以下、アイテム)を販売する小売店において、在庫管理や機会損失回避の観点から、日々変化する気象条件などの外部条件に起因する需要変動を把握することが求められている。こうした顧客の需要変動は、一般的に外部条件に影響され、特に気象条件や曜日条件のような外部条件に強く影響されると考えられる。また、小売店の現場では、非常に多くのアイテムが存在し、需要変動の要因が類似したアイテムをグルーピングすることで、管理業務の効率化が望まれる。ここで、例えば「金曜日であり、かつ気温が低い日には鍋の具材の売上が伸びる」など、需要変動の要因間の交互作用の存在も考えられる。したがって、需要変動の要因における交互作用効果を考慮して、アイテムをクラスタリングする必要がある。

そこで本研究では、交互作用効果をモデル化可能な Factorization Machines [1](以下、FM)を導入し、FMにより推定されたパラメータを活用して、需要変動要因の類似性によってアイテムのクラスタリングを行う手法を提案する。具体的には、FMで推定された各アイテムのパラメータに対して、多次元のデータを複数の正規分布の混合によって表現する Gaussian Mixture Model[2](以下、GMM)を適用し、クラスタリングを行う。特定の要因に著しい影響を受けているアイテムについては、主成分分析を用いて外れ値として効率的に検出する。最後に、実データに提案手法を適用し、得られた結果に対して考察を与え、提案手法の有効性を示す。

2 準備

2.1 対象問題

本研究では、某小売店チェーン店舗 A において、生鮮カテゴリに属するアイテムの購買履歴データを対象とする。生鮮カテゴリのアイテムの多くは、店舗ごとに加工方法を変えるなど、販売方法を柔軟に変更可能なため、需要変動要因を把握する重要性が高いアイテムである。需要変動要因としては、消費者の生活や行動と深い関係のある気象条件、月、曜日、祝日といった外部条件を考慮する。これらの要因は店舗にて把握可能であり、その日の条件に合わせた販売方法や在庫管理を考えられる。さらに、各要因単独の効果だけでなく、それらの要因間の交互作用を考慮することで、より詳細な外部条件によるアイテムのグルーピングが可能である。

2.2 Factorization Machines(FM)

FMでは特徴間の交互作用を考慮可能であり、高い予測精度を示すことが一般に知られている。外部条件を表す説明変数ベクトルを $\mathbf{x} = (x_1, x_2, \dots, x_I)^T$ 、目的変数を y 、バイアス項を w_0 、重みベクトルを $\mathbf{w} = (w_1, w_2, \dots, w_I)^T$ とする。

ここで、特徴量間の交互作用を表す交互作用行列は、 K 次元ベクトル $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{iK})^T$ を要素とする $I \times K$ の行列 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_I]^T$ で表せる。 $K \ll I$ のとき、FMのモデル式は式(1)で表される。

$$y = w_0 + \sum_{i=1}^I w_i x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^I \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \quad (1)$$

特徴量間の交互作用を表す右辺第三項の $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ は \mathbf{v}_i と \mathbf{v}_j の内積を表し、 $\frac{1}{2}I(I-1)$ 個の内積値が存在する。ここで内積値の絶対値が大きいほど、 i 番目と j 番目の特徴間に強い交互作用があると解釈できる。本研究では FM の学習によって推定されるパラメータ $w_0, \mathbf{w}, \mathbf{V}$ を用いたクラスタリングを行う。

3 提案手法

3.1 提案への着想と概要

本研究では、外部条件の交互作用を考慮しアイテムをグルーピングすることを目指す。そこで、FMを用いて各アイテムの売上数量を、外部条件とその交互作用を用いて表現し、推定されたパラメータからアイテムをクラスタリングする手法を提案する。具体的には、FMで推定された多数のパラメータに対して GMM を適用する。GMM は、複雑な構造をもつ多次元のデータを複数の正規分布の混合によって表現可能な汎用性のあるモデルであり、結果からクラスタリングが可能となる。ただし、特定の要因が著しく需要に影響しているアイテムについては、推定されたパラメータ値が非常に大きな値となるため、クラスタリングの際にクラスサイズに大きな偏りが生じてしまう。そこで、推定された FM の各パラメータについて事前に主成分分析を行い、主成分得点の分布から需要変動に強い影響を受けるアイテムを検出し、その要因を抽出する。

3.2 提案アルゴリズム

いま、分析対象として M アイテムに対して、 N 日分の売上データを考える。第 n 日の i 番目の外部条件を、 $x_{ni} \in \mathbb{R}$ 、外部条件を表す説明変数ベクトルを $\mathbf{x}^n = (x_{n1}, x_{n2}, \dots, x_{nI})^T$ 、日ごとのアイテム m の売上数量を表す目的変数を $y_n^m \in \mathbb{R}$ とする。また、FM のバイアス項を w_0^m 、重みベクトルを $\mathbf{w}^m = (w_1^m, w_2^m, \dots, w_I^m)^T$ 、交互作用の重みベクトルを $\mathbf{u}^m = (u_{1,2}^m, u_{1,3}^m, \dots, u_{I-1,I}^m)^T$ とする。ただし、 $m = 1, 2, \dots, M$ 、 $n = 1, 2, \dots, N$ である。以上のもとで提案アルゴリズムは以下の 5 つのステップから成る。

Step1) FM を用いた各アイテムの売上の予測

外部条件ベクトル \mathbf{x}_n からアイテム m の売上 y_n^m を予測する FM モデルを構築し、各アイテムについて各要因の売上への寄与度を表すパラメータ $w_0^m, \mathbf{w}^m, \mathbf{u}^m$ を得る。

Step2) パラメータを用いた主成分分析

*早稲田大学
†日本気象協会

得られた w_0^m , w^m 及び u^m を各次元とするベクトル空間上の各アイテムの推定したパラメータに対して主成分分析を適用する。

Step3) 主成分得点による、特定の要因に著しく強い影響を受けているアイテムの検出及びクラスタリング

各主成分において、主成分得点の標準偏差を σ として主成分得点の絶対値が $\alpha\sigma$ 以上のアイテムを検出する。なお、 α は外れ度の程度を示す値であり、適切に設定する。そのうえで因子負荷量により、各主成分(クラス)に寄与する要因を特定し、それを元に検出したアイテムのクラスに要因のラベル付けをする。

Step4) 残ったアイテムのクラスタリング

Step3 で検出されなかったアイテムについて、 w_0^m , w^m 及び u^m を各次元とするベクトル空間で GMM を適用する。GMM において、各アイテムは所属確率が最大となる潜在クラスにのみ所属すると仮定する。

Step5) クラスタリング結果の解釈

各クラスの平均ベクトルを用いて、各クラスに所属するアイテムの売上に寄与する要因のラベル付けを行い、解釈を行う。

4 実データ分析

4.1 分析条件

本分析では、某小売チェーン店舗 A において販売される生鮮カテゴリに属するアイテムの購買履歴データを対象とする。FM を適用する際の分析対象期間は、2013 年 1 月 8 日から 12 月 31 日 ($N = 358$) までとした。販売されない時期があるアイテムを除くため、分析対象アイテムは、対象期間内で連続して 1 週間以上売上のないアイテムを除外した 209 種 ($M = 209$) とした。FM における交互作用の低次元パラメータ数は結果の解釈性を考慮し $K = 10$ とした。目的変数は、各アイテムの日々の売上数量を対象期間の売上数量の最高値で割った値とし、説明変数は、各日の気象条件と、曜日、月、祝日の One Hot ベクトルとする。気象条件として用いた変数を表 1 に示す。ただし、気象条件も目的変数と同様に正規化する。

表 1: FM の説明変数 (量的変数)

特徴量	算出方法
気温前日差 (°C)	(当日の平均気温) - (前日の平均気温)
気温 3 日前差 (°C)	(当日の平均気温) - (3 日前の平均気温)
気温 7 日前差 (°C)	(当日の平均気温) - (7 日前の平均気温)
日射量前日差 (W/m ²)	(当日の平均日射量) - (前日の平均日射量)
日射量 3 日前差 (W/m ²)	(当日の平均日射量) - (3 日前の平均日射量)
日射量 7 日前差 (W/m ²)	(当日の平均日射量) - (7 日前の平均日射量)
相対湿度 (%)	当日の平均相対湿度
降水量 (mm)	当日の合計降水量

また、Step3 の主成分分析における主成分の数は、累積寄与率が 0.8 以上となる 50、 α は $\alpha = 3.0$ と設定した。GMM におけるクラス数は結果の解釈性を考慮し、4 とした。

4.2 主成分分析で検出されたアイテムの分析

提案手法の Step3 で外れ値として検出されたアイテムを表 2 に示す。「主な需要変動要因」は、各主成分で最大の因子負荷量を持つ要因を示す。また、抜粋したアイテムは主成分得点が正の方向に大きいため、売上数量に与える影響は正である。

表 2: 主成分分析で検出されたアイテムの分析結果

主成分	アイテム (一部抜粋)	主な需要変動要因
第 4 主成分	ますの押し寿司	4 月と水曜日の交互作用
第 5 主成分	焼きそばかつ丼セット	2 月と水曜日の交互作用
第 11 主成分	牛豚ミンチ, 3 種のソーセージ	日射量前日差と 5 月の交互作用

表 2 から、例えば、ますの押し寿司は 4 月の水曜日、牛豚ミンチ、3 種のソーセージは日射量が前日に比べて多い 5 月に著しく売上が伸びる傾向にあると解釈できる。各アイテム群に対して、該当する外部条件の日に入荷数量を増やすことで機会損失を防ぐことが可能であると考えられる。

4.3 主成分分析で検出されなかったアイテムの分析

提案手法の Step3 で検出されなかった 150 個のアイテムに対して GMM を適用した結果を表 3 に示す。ただし、表 3 には各クラスに所属するアイテムの平均ベクトルから距離が近いアイテムを抜粋したものと、需要変動への影響が最大の要因を示している。また、「符号」は要因が売上に与える影響の正負である。「所属数」は各クラスへの所属アイテムの個数を示している。

表 3: GMM によるアイテムクラスタリングの分析結果

クラス	アイテム (一部抜粋)	主な需要変動要因	所属数	符号
1	れんこん天, イカフライ	木曜日と 2 月の交互作用	14	+
2	ハム, ウィナー	土曜日と 5 月の交互作用	49	+
3	肉団子, あらびきポーク	気温 3 日前差と祝日の交互作用	13	-
4	さけ刺身, 刺身盛り合せ	気温前日差	74	+

表 3 より、クラス 1 に所属するアイテムは 2 月の木曜日に、クラス 2 に所属するアイテムは 5 月の土曜日に売上が伸びると解釈できる。また、クラス 4 に所属するアイテムは、気温が前日に比べて上昇する日には売上が増加すると解釈できる。一方、クラス 3 に所属するアイテムは、主な需要変動要因である「気温 3 日前差と祝日の交互作用」の符号が負であることから、3 日前に比べて気温が低い祝日には売上が減少すると解釈することができる。

以上より、提案手法を用いてアイテムの需要変動要因を効率的に把握可能であることが示された。これにより、外部条件に基づいた、効率的なアイテムの加工方法の変更や仕入れ量の調整、機会損失の削減が期待される。

5 まとめと今後の課題

本研究では、生鮮食品を扱う小売店の購買履歴データを対象とし、気象条件や曜日条件といった外部条件及びその交互作用を考慮してアイテムの需要変動要因を効率的に把握可能な分析手法を提案した。また、実データに適用することで提案手法の有効性を示した。

今後の課題としては、割引率の考慮、他店舗を加味した分析、FM や GMM のハイパーパラメータの定量的な決定が挙げられる。

参考文献

- [1] S. Rendle, "Factorization Machines," *Proc. 2010 IEEE International Conference on Data Mining*, 2010
- [2] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.