

非負値行列因子分解に基づく気象条件と商品売上パターンの 関係分析モデルに関する一考察

岡山 成[†] 山下 遥[‡] 三川 健太[§] 後藤 正幸[¶] 吉開 朋弘^{||}
早稲田大学[†] 早稲田大学[‡] 湘南工科大学[§] 早稲田大学[¶] 日本気象協会^{||}

1. 研究背景と目的

近年、多くの小売店は様々なマーケティング施策を立案するために、購買履歴データを分析して、顧客の嗜好や商品アイテムの特徴の把握を試みている。ここで、小売店で販売されている様々な商品アイテムの売上と気象要素の関連性に着目すると、天候などの気象条件がアイスクリームやビールといった気象感応度が高い商品の売上に強い影響を与えると考えられる。このような商品の品揃えや在庫管理に結び付けるためには、気象要因を考慮した分析が望まれる。

そこで岡山ら [1] は気象条件と商品の売上パターンの関係を分析するモデルとして、購買日時とアイテムの売上数量のデータを用いて、日付ごとの商品アイテムの売上個数を要素とする行列を作成し、この行列表現されたデータの中に潜在するパターンを非負値行列因子分解 (以下, NMF) により抽出する方法を提案した。これにより、各クラスターの代表的な商品アイテムを求め、日付のクラスターに気象データを結びつけることで商品アイテムのクラスターの特徴を解釈し、気象条件から生じる購買パターンを明らかにすることが可能となる。

しかし、従来研究 [1] では気象データの季節効果と曜日効果が分離されていないため気象感応度が高い商品のクラスターリングにこれらの要因が強い影響を与えてしまうという課題があった。そこで本研究では、日付・商品アイテム行列から気象データの季節効果と曜日効果を除去し、その行列に対して従来手法 [1] を適用することで気象感応度が高い商品の抽出を試みる。また、馬賀ら [2] が作成した Tweet データに基づく体感気温を用い、より気象感応度が高い商品の抽出方法について検討する。加えて実データの分析を通じ、提案する分析モデルの有効性を示す。

2. NMF による気象条件-商品売上パターン分析

NMF は観測行列 A を二つの行列 U, V へと分解し、パターン抽出を行う手法であり、岡山ら [1] は日付・商品カテゴリ行列に対して NMF を適用し、得られた行列 U, V の双方に着目した。ここで、日付の集合を $\mathcal{I} = \{I_i : 1 \leq i \leq N\}$ 、商品アイテムの集合を $\mathcal{J} = \{J_j : 1 \leq j \leq M\}$ と定義する。日付 I_i に対し、商品アイテム J_j が売れた個数を a_{ij} と表し、日付・商品アイテム行列を $A = [a_{ij}] \in \mathcal{R}^{N \times M}$ と定義する。 L 個の潜在クラスターの集合 $\mathcal{K} = \{K_k : 1 \leq k \leq L\}$ の寄与度を要素に持つ行列を $U = [u_{ik}] \in \mathcal{R}^{N \times L}$ とし、商品アイテム J_j のクラスター K_k の寄与度を要素にもつ行列を $V = [v_{kj}] \in \mathcal{R}^{L \times M}$ と定義する。

まず、行列 V を用いて各商品アイテムのパターン抽出を行う。各商品アイテム J_j がどのクラスターに属するかを各商品アイテム J_j が持つ L 個のクラスターの寄与度 v_{kj} のうち、最も高い v_{kj} のクラスター K_k をその商品アイテム J_j の所属先として決定する。次に、日付 I_i のクラスター K_k の寄与度 u_{ik} から、上記の商品アイテム J_j の考え方と同様に各日付 I_i が属するクラスターを決定し、寄与度 u_{ik} と各日付 I_i の気象要素 W_{i1} (平均気温), W_{i2} (体感気温), W_{i3} (相対湿度), W_{i4} (降水量), W_{i5} (日射量) を用いることで、気象要素 W_{iw} のクラスター K_k の重み付き平均 h_{wk} を算出することができる。なお、 $\delta(I_i, K_k)$ は日付 I_i がクラスター K_k に属するとき 1、そうでない場合 0 となるインジケータ関数である。

$$h_{wk} = \sum_{1 \leq i \leq N} \left(\frac{u_{ik} \delta(I_i, K_k)}{\sum_{1 \leq i' \leq N} u_{i'k} \delta(I_{i'}, K_k)} * W_{iw} \right) \quad (1)$$

式 (1) で得た重み付き平均 h_{wk} より、各商品アイテムが所属するクラスターの意味付けを行うことができる。

3. 分析手法

従来研究 [1] では、気象データの季節効果と曜日効果が各商品アイテムのクラスターに強い影響を与えているため、気象感応度が高い商品の抽出が困難であるという問題が存在する。そこで、日付・商品アイテム行列から気象データの季節効果と曜日効果の除去を行う方法を以下に示す。

3.1 気象データの季節効果の除去

気象データの季節効果の除去を行うために日付・商品アイテム行列に移動平均法を適用する。移動平均法とは時系列データにおいて一定区間の平均を取り、データを平滑化する手法であり、時系列の中長期的な傾向を把握することができる。また、気象データの季節効果を除去した日付・商品アイテム行列を $A' = [a'_{ij}] \in \mathcal{R}^{N \times M}$ と定義する。

3.2 曜日効果の除去

一般的に曜日ごとに商品の売上傾向は異なるため、本研究では曜日ごとの売上の変動傾向を曜日効果と定義する。まず、日付 I_i の商品アイテム J_j の売上の平均からの差分を要素にもつ行列 $B = [b_{ij}] \in \mathcal{R}^{N \times M}$ と定義する。ただし、 b_{ij} は式 (2) に従うものとする。

$$b_{ij} = a'_{ij} - \sum_{1 \leq i \leq N} \frac{a'_{ij}}{N} \quad (2)$$

表 1: 商品アイテムに着目した各クラス (上位 4 点のみ)

	クラス 1	クラス 2	クラス 3	クラス 4	クラス 5	クラス 6	クラス 7	クラス 8	クラス 9	クラス 10
	小松菜 A	にんじん A	バラ肉 A	白菜 A	にんじん B	煮豚	えのき	ねぎ A	みかん A	ほうれん草 A
	ねぎ B	いなり	りんご	レタス	みかん B	ちんげん菜	ロース A	イチゴ	サーモン	えりんぎ
	れんこん	さんま	ロース B	ほうれん草 B	ほうれん草 C	ステーキ	ロース C	小松菜 B	寿司 A	ぶどう
	ねぎ C	トマト	ハム A	寿司 B	もも肉	わさび	みかん C	ロース D	寿司 C	みかん D
平均気温 h_{1k} [°C]	15.69	23.34	18.64	14.41	17.89	9.98	17.19	8.52	10.41	22.52
体感気温 h_{2k} [°C]	12.65	20.03	14.35	7.68	14.12	8.29	13.42	9.70	7.53	17.42
相対湿度 h_{3k} [%]	68.23	67.52	65.53	62.80	72.80	67.30	71.09	62.27	64.58	65.10
降水量 h_{4k} [mm]	2.41	8.76	4.12	1.01	13.02	3.04	6.72	2.47	1.39	6.94
日射量 h_{5k} [MJ/m ²]	9.88	11.22	11.23	9.09	8.56	7.86	9.40	8.94	9.85	13.86

表 2: 日付に着目した各クラス (上位 4 点のみ)

クラス 1	クラス 2	クラス 3	クラス 4	クラス 5	クラス 6	クラス 7	クラス 8	クラス 9	クラス 10
2013/11/6	2013/10/9	2014/10/19	2014/11/18	2013/10/23	2014/12/17	2014/11/5	2013/12/8	2013/11/20	2014/10/9
2013/11/5	2013/10/8	2014/10/20	2014/11/17	2013/10/22	2014/12/18	2014/10/28	2013/12/7	2013/11/23	2014/10/10
2013/11/2	2013/10/5	2014/10/21	2014/11/16	2013/10/26	2014/12/16	2014/10/29	2013/12/9	2013/11/27	2014/10/8
2013/11/3	2013/10/6	2014/10/17	2014/11/19	2013/10/24	2014/12/15	2014/11/2	2013/12/6	2013/11/22	2014/10/11

次に、式 (2) で求めた b_{ij} を用いて $l(i): i \rightarrow \{1, 2, \dots, 7\}$ (ただし、1 は月曜、2 は火曜、..., 7 は日曜と表す) 曜日の商品アイテム J_j の平均売上効果 (曜日効果) を算出する。また、その値を要素にもつ行列を $D = [d_{kj}] \in \mathcal{R}^{7 \times M}$ と定義する。ただし、 d_{kj} は式 (3) に従うものとする。

$$d_{kj} = \sum_{i:l(i)=k} \frac{b_{ij}}{n(l(i))} \quad (3)$$

そして、式 (3) で求めた d_{kj} を用いて曜日効果 d_{kj} を除去した日付 I_i の商品 J_j の売上個数を算出する。また、その値を要素に持つ行列を $C = [c_{ij}] \in \mathcal{R}^{N \times M}$ と定義する。ただし $n(k)$ は日付の集合 I のうち曜日効果が k の値に該当する要素の合計個数とし、 c_{ij} は式 (4) に従うものとする。

$$c_{ij} = a'_{ij} - d_{l(i)j} \quad (4)$$

最後に a_{ij} の代わりに、式 (4) より求めた c_{ij} を 2 節の従来手法に適用し、本提案手法の有効性を検証する。

4. 分析

4.1 分析データ

本研究では中部地方の某小売店の購買履歴データを分析対象とする。分析手法の有効性を検討するため、代表店舗 (愛知県) の 2013 年、2014 年の 10 月 1 日から 12 月 21 日の生鮮カテゴリデータを 3 節で提案した分析手法を用いて分析する。日付数は $N=164$ 、商品アイテム数は $M=6948$ であり、事前分析により、クラスタ数を $L=10$ と設定した。また、日付・商品アイテム行列の要素 a'_{ij} は、7 日間の移動平均法によって求めた。

4.2 分析結果と考察

分析の結果として、商品アイテムと日付のクラスターリングの結果を表 1、表 2 に示す。

商品アイテムに着目したクラスタより、平均気温 h_{1k} が高いクラスタ 2, 10 は暑い日に多く売れる特徴を持つ商品クラスタであると解釈できる。一方で平均気温 h_{1k} が低いクラスタ 6, 8 は寒い日に多く売れるアイテムのクラスタという解釈を行うことができる。

また日付に着目したクラスタより、2014 年に多く売れる特徴を持つ商品アイテムのクラスタと 2013 年に多く売れる特徴を持つ商品アイテムのクラスタに分かれる傾向にあることが確認された。さらに 10 月から 12 月のそれぞれの月に多く売れる特徴を持つ商品アイテムのクラスタに分かれる傾向があることも確認された。

ここで、表 1 における平均気温 h_{1k} 以外の気象要素に着目すると、降水量 h_{4k} が多いクラスタ 5 は雨の日に多く売れる特徴を持つ商品クラスタであると解釈できる。また体感気温 h_{2k} は平均気温 h_{1k} と比べて低くなる傾向にある。10 月から 12 月の秋季の購買履歴データを用いているため一般に体感気温は通常の気温と比べて低くなる。しかし、クラスタ 8 の体感気温 h_{2k} は平均気温 h_{1k} と比べて高く、日付のクラスタから冬に多く売れるという特徴を指摘することができる。よって、クラスタ 8 に所属する商品アイテムは、冬の中で体感的に暑いと感じる日に多く売れる特徴を持つクラスタであると解釈でき、気象感応度が高い商品アイテムが所属しているクラスタであると考えられる。以上のように、購買履歴データに本提案モデルを適用することによって、気象感応度が高い商品の抽出を行うことができる。

5. まとめと今後の課題

本研究では従来研究 [1] において考慮されていなかった気象データの季節効果と曜日効果の除去を行うことにより気象感応度が高い商品の抽出を行うことができた。今後の課題として、NMF を拡張させた手法やその他の手法の適用が挙げられる。これにより、より気象感応度が高い商品の抽出を行えることが期待される。

参考文献

- [1] 岡山 成, 山下 遙, 後藤正幸, 吉開朋弘 “気象条件を考慮した商品売上パターン分析に関する一考察,” 日本経営工学会 2016 年秋季大会予稿集, pp. 42-43, 2016.
- [2] 馬賀嵩士, 三川健太, 後藤正幸, 吉開朋弘, “気象情報と Tweet データの統合的分析による体感気温の定量化に関する一考察,” 第 39 回情報理論とその応用シンポジウム (SITA2016), pp. 567-572, 2016.