

# 非負値テンソル因子分解に基づく気象条件と購買パターンの 関係分析モデルとその需要予測への応用に関する研究

情報数理応用研究

5216C012-1 岡山成  
指導教員 後藤正幸

## Relational Analysis Model of Weather Conditions and Sales Patterns Based on Nonnegative Tensor Factorization and Its Application to Demand Forecasting

OKAYAMA Sei

### 1 研究背景・目的

多様な商品を所在地の異なる多店舗で販売する小売チェーンにおいて、気象条件の変化に伴う需要変動に起因する在庫過多や廃棄処分が課題となっている。これに対し、蓄積された多様な大規模データを分析することで、ミクロな観点からのより細かな在庫管理を可能とすることが期待されている。すなわち、商品の品揃えや在庫管理の問題をデータ分析に基づいて解決しようとする場合、気象条件を考慮した分析が望まれる。

このような議論のもと、著者ら [1] はすでに気象条件とアイテムの売上傾向を分析するために NMF (Non-negative matrix factorization) [2] を用いた分析モデルを提案している [1]。この手法では、“日付”、“アイテム”の2因子から構成される行列を二つの低次元な行列へと分解する。分解された行列のうちの一方は日付の特徴行列を表し、もう一方はアイテムの特徴行列を表す。この日付の特徴行列と気象条件の情報を結びつけることで、気象条件から生じる売上パターンを抽出することが可能である。さらに得られた行列を用いて、各クラスターの代表的な商品を把握し、気象条件に影響されやすい商品を明らかにすることができると期待される。

しかし、従来研究 [1] の手法では日付とアイテムの組み合わせに対して、1店舗のみの購買データしか用いることができず、複数の店舗や地域の売上を同時に分析することができない。実際には店舗や地域によって気象効果は差異があると考えられるため、複数の店舗や地域の差異を同時に分析できるモデルが望まれる。

そこで本研究では、多店舗の購買データから“日付”、“アイテム”、“店舗”の三次元配列で表されたテンソルを形成し、NTF (Non-negative tensor factorization) [3] を適用することで、複数の店舗において共通して気象感応度が高い商品を抽出可能な分析モデルを提案する。さらに、多店舗の購買データを圧縮した店舗情報を説明変数として、重回帰分析を用いることで、需要の予測精度の向上を図るとともに、より詳細な気象要素を用いた分析を行う。

### 2 NMF を用いた気象要因の抽出 [1]

NMF を用い、アイテムの売上日時とアイテムごとの売上数量から形成される行列  $A$  を二つの行列  $U, V$  の積で式 (1) のように近似することで、アイテムと日付の特徴を抽出し、日付の情報を基に気象の情報と結びつけることが可能となる。このとき、行列  $U$  は日付の複数の特徴を表すことができ、日付とクラスター (特徴) の組み合わせによって形成される行列である。同様に、行列  $V$  はアイテムの複数の特徴を表すことができ、アイテムとクラスター (特徴) の組み合わせによって形成される行列である。

$$A \approx UV^T \quad (1)$$

ここで、日付の集合を  $\mathcal{I} = \{I_i : 1 \leq i \leq N\}$ 、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、アイテム  $J_j$  が売れた個数を  $a_{ij}$  と表し、日付、アイテム行列を  $A = [a_{ij}] \in \mathcal{R}_+^{N \times M}$  と定義する。ただし、

$\mathcal{R}_+^{N \times M}$  は、非負のみの実数値を要素にもつ  $N \times M$  行列の集合である。また、日付のジャンルなどを示す  $L$  個のクラスターの集合  $\mathcal{K} = \{K_k : 1 \leq k \leq L\}$  の寄与度を要素に持つ行列を  $U = [u_{ik}] \in \mathcal{R}_+^{N \times L}$  とし、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  $V = [v_{jk}] \in \mathcal{R}_+^{M \times L}$  と定義する。

NMF では

$$\begin{aligned} & \underset{U, V}{\text{minimize}} \|A - UV^T\|_F^2 \\ & \text{subject to } \forall u_{ik} \geq 0, \forall v_{jk} \geq 0 \end{aligned} \quad (2)$$

の最適化問題を解くことで非負行列  $U, V$  へ分解する。ただし、 $\|\cdot\|_F$  は行列のフロベニウスノルムを表すものとする。

この最適解は式 (3), (4) を収束するまで繰り返すことにより、得ることができると期待される。ただし、 $t$  回目の更新により得られた  $U^{(t)}, V^{(t)}$  の各成分を  $u_{ik}^{(t)}, v_{jk}^{(t)}$  と定義する。

$$u_{ik}^{(t)} = u_{ik}^{(t-1)} \frac{(A^{(t-1)}V^{(t-1)})_{ik}}{(U^{(t-1)}V^{(t-1)T}V^{(t-1)})_{ik}} \quad (3)$$

$$v_{jk}^{(t)} = v_{jk}^{(t-1)} \frac{(U^{(t-1)T}A^{(t-1)})_{jk}}{(U^{(t-1)T}U^{(t-1)}V^{(t-1)T})_{jk}} \quad (4)$$

しかし、以上の手法では1店舗の購買データのみを対象とすることしかできない。すなわち、複数の店舗の購買傾向やアイテムの特徴などを同時に分析することは想定されていないため、店舗や地域による気象効果の差異を考慮した分析が行えないという課題がある。そこで、より多角的な分析の実施のためには多店舗の売上データを用いた分析モデルが期待される。

### 3 気象条件から生じる購買パターンの抽出

#### 3.1 Non-negative tensor factorization (NTF)

##### を用いた気象要因の抽出 [3]

NTF は NMF をより一般化した手法で、観測テンソル  $B$  を三つの行列  $U, V, Y$  のテンソル積で近似し、パターン抽出を行う手法である。NMF では2因子のクラスタリングが限界であったが、NTF では、テンソルの階数 (データの因子数) を増やすことで複数の因子をもつ行列に分解することができる。また、分解された行列の値をもとに因子をクラスタリングすることも可能とする。ここで、従来研究で行っていた“日付”、“アイテム”の2因子に新たな因子として“店舗”を追加することで、1店舗に依存した分析から複数の店舗において共通した分析へと拡張することができる。これにより、複数の店舗や地域による気象効果の差異などといった要素も把握できることが期待され、幅広い考察を行うことができる。

ここで、店舗の集合を  $\mathcal{S} = \{S_s : 1 \leq s \leq P\}$  と定義する。店舗  $S_s$  で日付  $I_i$  にアイテム  $J_j$  が売れた個数を  $b_{ijs}$  と表し、日付、アイテム、店舗のテンソルを  $B = [b_{ijs}] \in \mathcal{R}_+^{N \times M \times P}$  と定義する。ただし、 $\mathcal{R}_+^{N \times M \times P}$

は、非負のみの実数値を要素にもつ  $N \times M \times P$  からなるテンソルの集合である。また、店舗  $S_s$  のクラスタ  $K_k$  への寄与度を要素にもつ行列を  $Y = [y_{sk}] \in \mathcal{R}_+^{P \times L}$  と定義する。

NTF では

$$\begin{aligned} & \underset{U, V, Y}{\text{minimize}} \|B - U \otimes V \otimes Y\|_F^2 \\ & \text{subject to } \forall u_{ik} \geq 0, \forall v_{jk} \geq 0, \forall y_{sk} \geq 0 \end{aligned} \quad (5)$$

の最適化問題を解くことで非負行列  $U, V, Y$  へ分解する。ただし、 $\otimes$  はテンソル積を表すものとする。

この最適解は式 (6)–(8) を収束するまで繰り返すことにより、得ることができる。ただし、 $t$  回目の更新に得られた  $U^{(t)}, V^{(t)}, Y^{(t)}$  の各成分を  $u_{ik}^{(t)}, v_{jk}^{(t)}, y_{sk}^{(t)}$  と定義する。

$$u_{ik}^{(t)} = u_{ik}^{(t-1)} \frac{(B_{(1)}^{(t-1)} Z^{(1)(t-1)})_{ik}}{(U^{(t-1)} Z^{(1)(t-1)T} Z^{(1)(t-1)})_{ik}} \quad (6)$$

$$v_{jk}^{(t)} = v_{jk}^{(t-1)} \frac{(B_{(2)}^{(t-1)} Z^{(2)(t-1)})_{jk}}{(V^{(t-1)} Z^{(2)(t-1)T} Z^{(2)(t-1)})_{jk}} \quad (7)$$

$$y_{sk}^{(t)} = y_{sk}^{(t-1)} \frac{(B_{(3)}^{(t-1)} Z^{(3)(t-1)})_{sk}}{(Y^{(t-1)} Z^{(3)(t-1)T} Z^{(3)(t-1)})_{sk}} \quad (8)$$

ただし、テンソル  $B$  の3つの添字がそれぞれ行の添字となるように、テンソルの要素を行列の形に並び替えたものを  $B_{(1)} \in \mathcal{R}_+^{N \times MP}$ ,  $B_{(2)} \in \mathcal{R}_+^{M \times NP}$ ,  $B_{(3)} \in \mathcal{R}_+^{P \times NM}$  としている。また、 $Z^{(1)} = V \otimes Y$ ,  $Z^{(2)} = U \otimes Y$ ,  $Z^{(3)} = U \otimes V$  である。

### 3.2 気象要素を用いた解釈

本節では、日付、アイテム、店舗テンソル  $B$  に対して、NTF によって得られた日付の特徴行列  $U$  に気象条件を関連付け、解釈する方法を提案する。

まず、商品の特徴行列  $V$  を用いて各アイテムのクラスタリングを行う。各アイテム  $J_j$  がどのクラスタに属するかについて、各アイテム  $J_j$  が持つ  $L$  個のクラスタの寄与度  $v_{jk}$  のうち、最も高い  $v_{jk}$  のクラスタ  $K_k$  を所属先として決定する。次に、日付  $I_i$  のクラスタ  $K_k$  の寄与度  $u_{ik}$  から、上記のアイテム  $J_j$  の考え方と同様に各日付  $I_i$  が属するクラスタを決定する。

ここで、平均気温、前日との気温差、相対湿度、降水量の4つの気象要素を対象とし、気象条件の集合を  $\mathcal{Q} = \{Q_q : 1 \leq q \leq 4\}$  と定義する。日付  $I_i$  のとき、気象条件  $Q_q$  の大きさを  $w_{iq}$  と表し、日付、気象条件の行列を  $W = [w_{iq}] \in \mathcal{R}^{N \times 4}$  と定義する。そして、得られた日付の特徴行列  $U$  と日付、気象条件の行列  $W$  を用いて、各商品クラスタへの気象要素を用いた解釈を行う。まず、得られた寄与度  $u_{ik}$  と各日付の気象条件  $w_{iq}$  を用いて、気象要素  $Q_q$  のクラスタ  $K_k$  の重み付き平均  $h_{qk}$  を算出する。

$$h_{qk} = \sum_{i=1}^N \left( w_{iq} \frac{u_{ik} \delta(I_i, K_k)}{\sum_{i'=1}^N u_{i'k} \delta(I_{i'}, K_k)} \right) \quad (9)$$

ここで、 $\delta(I_i, K_k)$  は日付  $I_i$  がクラスタ  $K_k$  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。

そして、式 (9) で得た重み付き平均  $h_{qk}$  の値より、各商品クラスタに対して気象要素を用いた解釈を行うことができる。

## 4 店舗情報の圧縮に基づく需要予測

需要予測の手法として、重回帰や ARIMA モデルが挙げられる。一般的に商品の売上個数を予測する際には、目的変数を売上個数とし、説明変数は気象、曜日、店舗などといった要素とする。しかしながら本研究で対象として

いる事例では、多くの店舗のデータが含まれるため、店舗をそのまま変数として説明変数に用いると、パラメータ数が多くなり、需要の予測精度が悪化する可能性がある。そこで、店舗情報を表す変数を圧縮することができればパラメータ数の増加を抑え、需要の予測精度が向上することが期待される。本研究では NTF によって得られる店舗の特徴行列を用いて、店舗情報の圧縮を行う。

また、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する。そして店舗情報を圧縮する際に、店舗の差が強調されたクラスタリングを行い、それらを圧縮した店舗情報として需要予測に用いる。このとき、各店舗の購買データには曜日効果が含まれており、クラスタリングを行う際に日付の特徴行列に曜日効果が影響する可能性があるため、購買データから曜日効果を除去した後に、NTF を適用することを考える。

まず、店舗  $S_s$  における日付  $I_i$  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テンソルを  $C = [c_{ijs}] \in \mathcal{R}^{N \times M \times P}$  と定義する。ただし、 $c_{ijs}$  は式 (10) に従うものとする。

$$c_{ijs} = b_{ijs} - \sum_{i=1}^N \frac{b_{ijs}}{N} \quad \forall s = 1, \dots, P \quad (10)$$

次に、式 (10) で求めた  $c_{ijs}$  を用いて、店舗  $S_s$  における  $l(i) \rightarrow \{1, 2, \dots, 7\}$  (ただし、1 は月曜、2 は火曜、...、7 は日曜を表す) 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テンソルを  $D = [d_{rjs}] \in \mathcal{R}^{7 \times M \times P}$  と定義する。ただし、 $d_{rjs}$  は式 (11) に従うものとする。

$$d_{rjs} = \sum_{i: l(i)=r} \frac{c_{ijs}}{n_s(l(i))} \quad \forall s = 1, \dots, P \quad (11)$$

ただし  $n_s(r)$  は店舗  $S_s$  における日付の集合  $\mathcal{I}$  のうち曜日効果が  $r$  の値に該当する要素の合計個数とする。そして、式 (11) で求めた曜日効果  $d_{rjs}$  を用いて曜日効果を除去した店舗  $S_s$  における日付  $I_i$  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テンソルを  $F = [f_{ijs}] \in \mathcal{R}^{N \times M \times P}$  と定義し、 $f_{ijs}$  は式 (12) に従うものとする。

$$f_{ijs} = b_{ijs} - d_{l(i)js} \quad \forall s = 1, \dots, P \quad (12)$$

最後に、 $b_{ijs}$  の代わりに、式 (12) より求めた  $f_{ijs}$  に対して NTF を適用する。

## 5 気象条件と購買パターンの関係分析

本章では、気象条件に影響されやすい商品グループを明らかにするために、購買データに NTF を適用し、分析を行う。しかし、本研究では多店舗の購買データを用いているため、店舗数が多いと、地域性がクラスタの解釈に強く影響し、気象条件を用いたクラスタの解釈が難しくなる。そこで、同じ気象条件であることを仮定できる1地域の購買データに絞って分析を行うことで、地域性の影響を弱め、気象条件を用いたクラスタの解釈を行う。

### 5.1 対象データ

本研究では中部地方の某小売店の購買データを分析対象とする。本章では、提案した手法の有効性を検討するため、代表店舗 (石川県) の2012, 2013, 2014年の10月1日から11月30日の生鮮カテゴリデータに対して、NTF による分析を行った。日付数は  $N=183$ 、アイテム数は  $M=8146$ 、店舗数  $P=12$  とし、 $b_{ijs}$  を要素に持つ日付、アイテム、店舗テンソル  $B$  を用いた。また、事前分析により気象条件の解釈性の高いクラスタ数を探索的に求め、クラスタ数を  $L=15$  と設定した。気象条件については、各期 (日付) の平均気温 [°C]、前日との気温差 [°C]、相対湿度 [%]、日合計降水量 [mm] を用いた。

表 1: 日付クラスタ (寄与度上位 5 日付, 代表的な 10 クラスタ)

クラスタ 1	クラスタ 2	クラスタ 3	クラスタ 4	クラスタ 5	クラスタ 6	クラスタ 7	クラスタ 8	クラスタ 9	クラスタ 10
2012/10/6	2014/11/23	2013/10/2	2014/10/28		2014/10/1	2013/11/24			2013/11/13
2012/10/7	2014/10/4	2013/10/4	2014/10/20		2012/10/3	2014/11/30			2013/11/6
2012/10/13	2014/10/8	2013/10/6	2014/11/5		2013/10/1	2013/11/23			2013/11/8
2012/10/8	2014/10/3	2013/10/3	2014/11/26		2013/10/31	2013/11/29			2013/10/25
2012/11/18	2014/10/2	2013/10/8	2014/11/14		2013/10/4	2013/11/16			2013/11/15

表 2: 商品クラスタ (寄与度上位 5 商品, 代表的な 10 クラスタ)

	クラスタ 1	クラスタ 2	クラスタ 3	クラスタ 4	クラスタ 5	クラスタ 6	クラスタ 7	クラスタ 8	クラスタ 9	クラスタ 10
	あさり	メロン	エリンギ	ゼリー	肉団子	マッシュルーム	レタス	お好み焼き	柿	ほうれん草
	いか	丼	豚肉 A	いよかん	豚肉 B	もつ煮込み	ゼリー	キンカン	アボカド	豚肉 E
	レーズン	サラダ	ごぼう	ホタテ	豚肉 C	揚げ物	みかん A	お好み焼き	バナナ	寿司 B
	かまぼこ	ぶどう	にんじん A	弁当 A	豚肉 D	トマト A	そばろ	弁当 B	りんご	寿司 C
	たけのこ	花	にんじん B	おこわ	牛肉	トマト B	さば	寿司 A	パブリカ	寿司 D
平均気温 $h_{1k}$ ( $^{\circ}\text{C}$ )	14.33	17.97	18.18	13.55		17.74	9.07			10.20
気温差 (前日) $h_{2k}$ ( $^{\circ}\text{C}$ )	-0.24	0.15	-0.33	-0.27		0.18	-0.91			-0.33
相対湿度 $h_{3k}$ (%)	68.39	76.05	73.04	72.23		72.11	76.49			74.67
日合計降水量 $h_{4k}$ (mm)	10.82	11.04	9.30	9.25		3.58	17.09			24.72

表 3: 店舗クラスタ (寄与度上位 2 店舗, 代表的な 10 クラスタ)

クラスタ 1	クラスタ 2	クラスタ 3	クラスタ 4	クラスタ 5	クラスタ 6	クラスタ 7	クラスタ 8	クラスタ 9	クラスタ 10
店舗 A			店舗 B	店舗 C			店舗 D	店舗 E	店舗 F
									店舗 G

## 5.2 代表的な購買パターンの解釈

ここでは, 対象データに対して, 本分析モデルを適用することで, 気象感応度が高い商品の抽出を行う。

分析の結果として, NTF より算出した日付, アイテム, 店舗のクラスタリングの結果を表 1 から表 3 に示し, 気象条件および, その他から得られる解釈を行う。また, 表 2, 3 中の空欄は, これ以上そのクラスタに属す日付や店舗がないことを示す。

### 5.2.1 気象条件から得られる解釈

表 1 の日付クラスタに着目すると, クラスタ 2, 4 は 2014 年, クラスタ 3, 10 は 2013 年, クラスタ 1 は 2012 年に多く売れる特徴を持つクラスタに分かれ, それぞれの年のみに多く売れるといった年効果が確認された。この年効果は気象条件を用いた商品クラスタの解釈に強く影響する可能性がある。すなわち, 気象要素がその年だけでなく, 毎年のように売上に強く影響を及ぼす商品を把握することができれば, 実際の施策により結びつけやすくなる。そこで, 本研究では年効果がないクラスタを対象とし, 気象条件を用いた解釈を行うこととする。

表 2 の商品クラスタに着目すると, 気温  $h_{1k}$  より年効果がないクラスタ 6 を「暖かい日に売れる商品」のクラスタであると解釈する。一方で, 年効果がないクラスタ 7 は「寒い日に多く売れる商品」のクラスタであると解釈できる。

加えて, 前日との気温差  $h_{2k}$  と降水量  $h_{4k}$  に着目すると, クラスタ 6 は気温差  $h_{2k}$  が高く, 降水量  $h_{4k}$  が少ないことから, かなり暖かい日に売れる商品クラスタであると解釈できる。同様にして, クラスタ 7 は前日との気温差  $h_{2k}$  が低く, 降水量  $h_{4k}$  が多いことから, かなり寒

い日に多く売れる商品クラスタであると解釈できる。

よって, クラスタ 6 に属する「マッシュルーム」, 「もつ煮込み」, 「揚げ物」といった商品は暖かくなると, 需要が上昇する商品であり, 一方で, クラスタ 7 に属する「レタス」, 「ゼリー」, 「みかん」といった商品は寒くなると, 需要が上昇する商品であることが確認できた。すなわち, 事前に天気予報を把握することで, これらの気象条件に影響されやすい商品の在庫過多や廃棄処分を未然に防ぐことができる。

### 5.2.2 その他の解釈

気象条件以外に着目すると, 表 3 の店舗クラスタから, 店舗の特徴を詳しく把握することができる。例えば, クラスタ 1 より, 店舗 A では 2012 年に「あさり」, 「いか」といった商品が多く売れたことを把握することができる。

以上のことから, 各店舗ごとに異なる背景を分析する必要がある。例えば, 店舗 A, B, F, G では, ある年に共通して多く売れた商品が存在することから, それらの店舗がその年に多く講じたマーケティング施策とその年に多く売れた商品との因果関係について検討すべきと考えられる。

一方で, 店舗 C では肉製品が多く売れるという特徴が抽出されている。同様に, 店舗 D, E においても売れ筋商品による特徴が得られた。すなわち, 店舗 C, D, E ではそれぞれ肉, 炭水化物, フルーツに関連した商品が多く売れることから, それらの店舗が毎年講じているマーケティング施策とこれらの商品との因果関係についても検討すべきと考えられる。

以上より, 本分析モデルを適用することで, 気象条件に影響されやすい商品を明らかにすることができた。

## 6 店舗情報を考慮した商品の需要予測

一般に、小売チェーンにおける商品の需要予測では店舗特徴を考慮した予測モデルを検討することが望ましい。しかし、店舗数が多数になる場合には、単純に店舗毎に予測モデルを構築すると、十分な予測精度が得られない可能性がある。そこで、NTF の出力結果を用いて店舗をクラスタリングすることで店舗情報を圧縮し、重回帰分析の説明変数として用いることで、需要の予測精度の向上を図る。

### 6.1 NTF による店舗情報の圧縮

本節では、店舗情報の圧縮を行うために、購買データに NTF を適用し、分析を行う。このとき、地域性の影響を強めたクラスタリングが望まれるため、5 章の分析とは異なり、全国の店舗を対象とし、店舗数を多く設定する。

#### 6.1.1 対象データ

以下では、5 章と同様に中部地方の某小売店の購買データを分析対象とする。本節では、提案した手法の有効性を検討するため、2012, 2013, 2014 年の 10 月 1 日から 11 月 30 日の生鮮カテゴリデータを用いる。日付数は  $N=183$ 、アイテム数は  $M=8146$ 、店舗数  $P=160$  とし、曜日効果を除去した  $f_{ijs}$  を要素に持つ日付、アイテム、店舗テンソル  $F$  を用いた。また、適切なクラスタ数は探索的に求め、 $L=10$  と設定した。

#### 6.1.2 店舗情報の圧縮

NTF より算出した店舗のクラスタリングの結果を表 4 に示す。ただし、ここでは店舗クラスタに属する店舗を、その店舗が存在する県で表す。

表 4: 店舗クラスタ (5 店舗, 代表的な 5 クラスタ)

クラスタ 1	クラスタ 2	クラスタ 3	クラスタ 4	クラスタ 5
石川県	岐阜県	静岡県	長野県	富山県
富山県	岐阜県	静岡県	長野県	富山県
富山県	岐阜県	静岡県	長野県	富山県
石川県	岐阜県	静岡県	長野県	愛知県
富山県	岐阜県	静岡県	長野県	三重県

表 4 より、クラスタ 1 は石川、富山県、クラスタ 2 は岐阜県、クラスタ 3 は静岡県、クラスタ 4 は長野県、クラスタ 5 は富山県の店舗が属しており、これらの店舗で多く売れる傾向にあると解釈することができる。すなわち、各地域ごとに店舗がクラスタリングされることが確認できた。

### 6.2 店舗情報を考慮した商品の需要予測

本節では、表 4 で示した NTF の出力結果から、店舗をクラスタリングし、所属するクラスタを表すダミー変数を作成した。そして、それらを購買予測モデルに説明変数として取り組むことにより、圧縮された店舗の違いを考慮した予測モデルを実現する。

本研究では対象期間のうち、2013 年 10 月 1 日から 2013 年 11 月 30 日の気象要素、および同期間の生鮮カテゴリの売上数量データを学習データ、2014 年 10 月 1 日から 2014 年 11 月 30 日の同データをテストデータとして予測実験を行なった。説明変数は気温、前日との気温差、相対湿度、降水量、曜日変数、店舗変数とする。ここで曜日変数には金曜日を基準として、その他の曜日を 1 で指定する 6 つの  $\{0, 1\}$  を取るダミー変数を用いた。店舗変数は 160 店舗を 10 のクラスタへと分割し、9 つのダミー変数で表すことにする。目的変数は商品の売上個数を対象とした。精度評価として平均絶対誤差率 (MAPE) を用いる。

### 6.2.1 商品の需要予測

店舗情報を圧縮した説明変数 (10 店舗クラスタ) を重回帰分析に用いた場合と圧縮しなかった場合 (160 店舗) での予測精度の比較を行なった。ここでは、5 章の表 2 で示した気象条件に影響されやすい商品グループ (クラスタ 6,7) のうち、売上数が最も多いマッシュルーム (クラスタ 6)、さば (クラスタ 7) の 2 商品の MAPE を表 5 に示す。

表 5. 気象条件に影響されやすい 2 商品の MAPE [%]

商品	10 店舗 (圧縮)	全 160 店舗
マッシュルーム	47.43	86.63
さば	9.13	9.80

表 5 より、店舗クラスタを構成することで、需要の予測精度が向上することが確認できた。

### 6.2.2 気象条件と商品の売れ行きの関係の分析

本節では、学習データに対して、重回帰分析を行った結果、得られた気温、前日との気温差が与える商品の売上個数 (目的変数) への影響の度合い (偏回帰係数) を示す。以下では、表 2 のクラスタ 7 に所属する商品のみを分析対象とする。これは、5 章より、クラスタ 7 に所属する商品は気温、前日との気温差の 2 要素が売上に影響していることが分かっているが、どちらの要素が売上に最も影響しているかの確認をする必要があるからである。得られた気温、前日との気温差のそれぞれの偏回帰係数に着目し、考察を行う。以下に気温差、気温の偏回帰係数が負に値が大きくなる商品を 4 つ、それぞれ表 6 に示す。

表 6. 寒い日に売れる生鮮商品

気温差が負の日に売上増加		気温が低い日に売上増加	
商品	偏回帰係数	商品	偏回帰係数
白ねぎ	-1.842	サラダ	-0.453
柿	-0.240	いちご A	-0.365
さば	-0.195	いちご B	-0.273
黄金いか	-0.194	いちご C	-0.251

表 6 より、気温が低い日に売れる商品はフルーツが多く、一方で前日との気温差が負の日に売れる商品はフルーツやその他の商品が確認された。

よって、気温が低い日に売れる商品はフルーツが多いことから、季節性の影響が強いと考えられる。一方で、前日との気温差が負の日に売れる商品は、季節性の影響があまりなく、気象感応度が高いと考えられる。

## 7 まとめと今後の課題

本研究では NTF を導入することで、1 地域で共通した気象感応度が高い商品を抽出する分析方法と店舗情報を考慮した需要予測方法の提案を行った。本分析では、従来研究 [1] では考慮されていなかった店舗の特徴を分析することができ、また店舗情報を圧縮することで予測精度が向上することが確認された。

今後の課題として、NTF を用いた際に気象感応度が高い商品を抽出することが可能な適切なクラスタ数の決定やその他の気象条件も用いたクラスタの解釈が挙げられる。

### 参考文献

- [1] 岡山成, 山下遥, 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” 日本経営工学会秋季大会予稿集, pp.42-43, 2016.
- [2] L. Daniel, and H. Seung, “Algorithms for non-negative matrix factorization,” NIPS'00 Proceeding of the 13th International Conference on Neural Information Processing Systems, pp.535-541, 2000.
- [3] 松林 達史, 幸島 匡宏, 林 亜紀, 澤田 宏, “非負値テンソル因子分解を用いた購買行動におけるブランド選択分析,” 人工知能学会論文誌, pp.713-720, 2015.