

平成29年度 NTT数理データシステム学生奨励賞 (Visual Mining Studio)

外的要因を加味した小売店の売上要因の分析

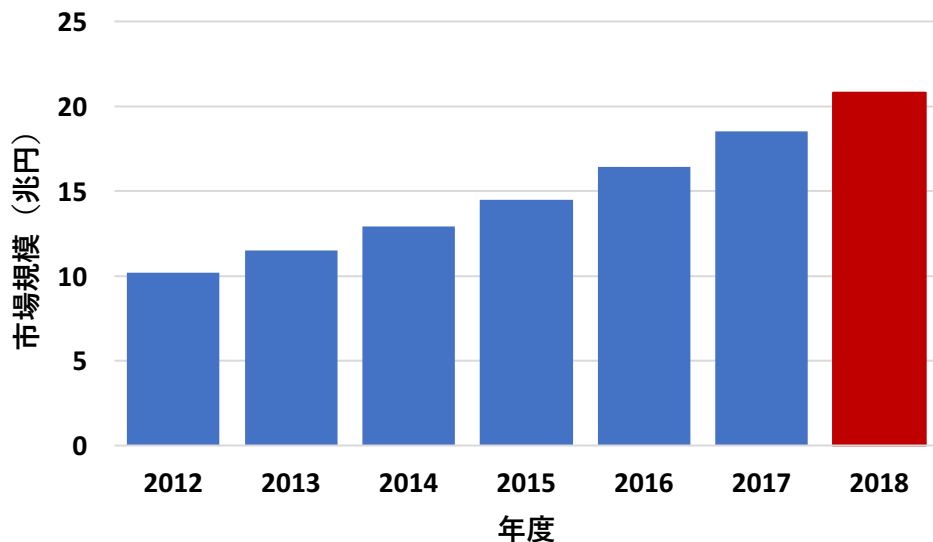
中央大学理工学研究科
佐藤由将

目次

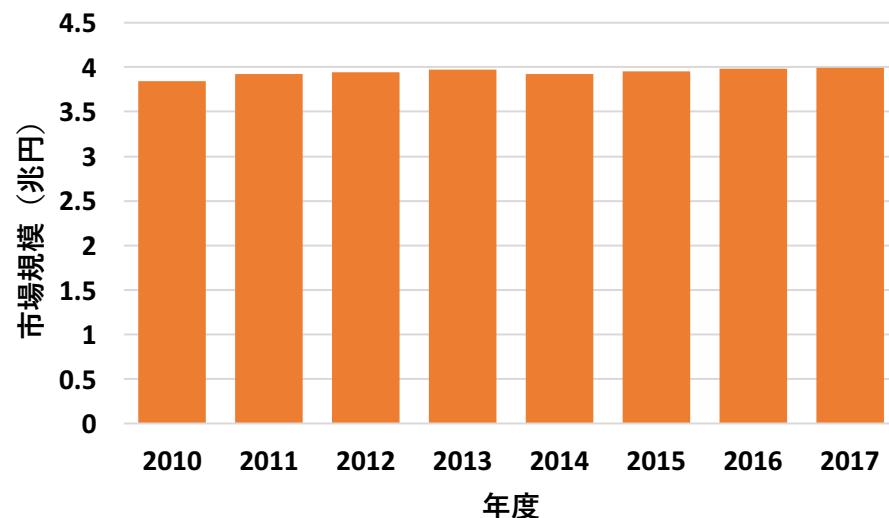
1. 背景
2. 目的
3. データ概要
4. 分析概要
 1. デシル分析
 2. 状態空間モデル
5. 分析結果
6. 考察
7. まとめと課題

背景

- 近年、インターネットの普及により、ウェブサイト上で商品を売買するECサイト市場が拡大の一途を辿っており、実店舗の経営状態を脅かすまでの存在となっている
- また、本研究で対象とするホームセンターの市場規模はオンラインショッピングなどの販売チャネルの多様化による競争激化、新規出店の際の費用高騰などにより停滞している



ECサイトの市場規模推移



ホームセンターの市場規模推移

目的

背景からの課題

- ホームセンター業界内ないしは同一チェーン内の経営戦略において、他店との差別化を図る必要性がある
- 上記の差別化は店舗毎の外的要因（天候、競合）の状況を考慮することを前提としているが、店舗内では商品カテゴリによって売上が異なることを考えると、カテゴリ毎に購買状況を分析することが重要である

目的

ホームセンターの店舗における商品カテゴリ毎の売上データを様々な要因に分解するモデルを構築し、その変数のパラメータの特徴を把握することで、商品発注などの店舗運営に有益な示唆を得ることを目的とする

データ概要

- ホームセンターチェーン店 1 店舗
- 立地：関東圏
- 使用データ：POSデータ
- データ期間：2016年6月5日～2017年5月31日
- 商品カテゴリ：29カテゴリ

分析概要

- 対象データ：部門毎の日別売上データ

データ分析の流れ

デシル分析 (VMS)

1. 各カテゴリ毎の年間売上個数を用いてデシルランクを割り当てる
2. デシルランク 1 を上位ランク、デシルランク 5 を中位ランク、デシルランク 9 を下位ランクとする

状態空間 モデル (R, Stan)

- 上記のランク毎に属するカテゴリの日毎の売上金額に対して、次ページで示す状態空間モデルを適用する

特徴把握 (VMS)

- 分解された要因を、カテゴリ間やランク間で比較し、特徴を把握する

デシル分析

目的：各カテゴリをランク付けすることで、状態空間モデルによる結果を各ランクに属するカテゴリで比較し、より精緻な考察を得るため

使用ソフトウェア：VMS

ランク	カテゴリ	デシルランク
1	daily	1
2	pet	1
3	food	1
4	hbc	2
5	metallic	2
6	dining/kitchen	2
7	stationary	3
8	plant	3



22	housing/exterior	8
23	furniture	8
24	audio visual	8
25	hobby	9
26	elect	9
27	sleep	9
28	heating/cooling	10
29	tenant pet	10

上位


- daily
- pet
- food

中位

- toilet
- work
- agri

下位*

- hobby
- elect
- sleep



*デシルランク10の商品の売上金額が、分析に使用する対象として不十分であると判断し、ランク9の商品を対象とした

状態空間モデル

目的：各カテゴリの売上金額を複数の要因に分解し、売上に寄与している要因を特定するため

使用ソフトウェア：R、Stan

売上金額 Y_t を以下の要因に分解する
(尚、以下に示すモデルは山口らのモデル^{*1}を参考にした)

$$Y_t = T_t + W_t + R_t + E_t + A_t + v_t$$

Y_t : 日毎の売上金額
 T_t : トレンド成分
 W_t : 曜日成分
 R_t : 天候成分
 E_t : イベント成分
 A_t : 自己回帰成分
 v_t : 残差

自己回帰成分

$$A_t = a_1 A_{t-1} + a_2 A_{t-2} + \varepsilon_t^A$$
$$\varepsilon_t^A \sim N(0, \sigma^A)$$

残差

$$v_t \sim N(0, \sigma)$$

*1 : p22の参考文献2を参照

状態空間モデル

トレンド成分

*2次の確率差分方程式を使用

$$T_t - T_{t-1} \approx T_{t-1} - T_{t-2}$$

$$T_t = 2T_{t-1} - T_{t-2} + \varepsilon_t^T \quad \varepsilon_t^T \sim N(0, \sigma^T)$$

曜日成分

$$W_t = b_t + h_t^1 s_{sun} (b_{sun,t} - b_t) \\ + h_t^2 \{s_{fri} (b_{fri,t} - b_t) + s_{sat} (b_{sat,t} - b_t)\}$$

第1項：週の基本項 $\sum_{l=1}^7 b_{t-l} \approx 0 \rightarrow b_t = -\sum_{l=1}^6 b_{t-l} + \varepsilon_t^W \quad \varepsilon_t^W \sim N(0, \sigma^W)$

第2項：祝日効果（直近の日曜の売上との類似度を表す） $0 \leq s_{sun} \leq 1$

第3項：祝日の前日効果
（直近の金曜と土曜の売上との類似度を表す） $0 \leq s_{fri}, s_{sat} \leq 1, 0 \leq s_{fri} + s_{sat} \leq 1$

	h_t^1	h_t^2
1	月～金の祝日	祝日でない月～木かつ翌日が祝日
0	それ以外	それ以外

状態空間モデル

天候成分

$$R_t = c_{rain} \times rain_val_t$$

* c_{rain} を推定し、
 $c_{rain} < 0 \rightarrow$ 売上減少
 $c_{rain} > 0 \rightarrow$ 売上増加

天気	$rain_val_t$
晴or曇	0
雨	0.5
大雨or雪	1

イベント成分

*分析対象の店舗は大型イベント施設の近隣であることから、競馬開催の売上への影響を表現する

$$E_t = c_{event_t} \times event_val_t$$

$$c_{event_t} = c_{event_{t-1}} + \varepsilon_t^E \quad \varepsilon_t^E \sim N(0, \sigma^E)$$

イベント	$event_val_t$
なし	0
あり	1

状態空間モデル

上記のモデルを状態ベクトル x_n を用いて

$$x_n = [x_{T,n}^t \mid x_{W,n}^t \mid x_{R,n}^t \mid x_{E,n}^t \mid x_{A,n}^t]^t$$

とすることで、状態空間モデル

$$x_n = F_n x_{n-1} + G_n v_n$$

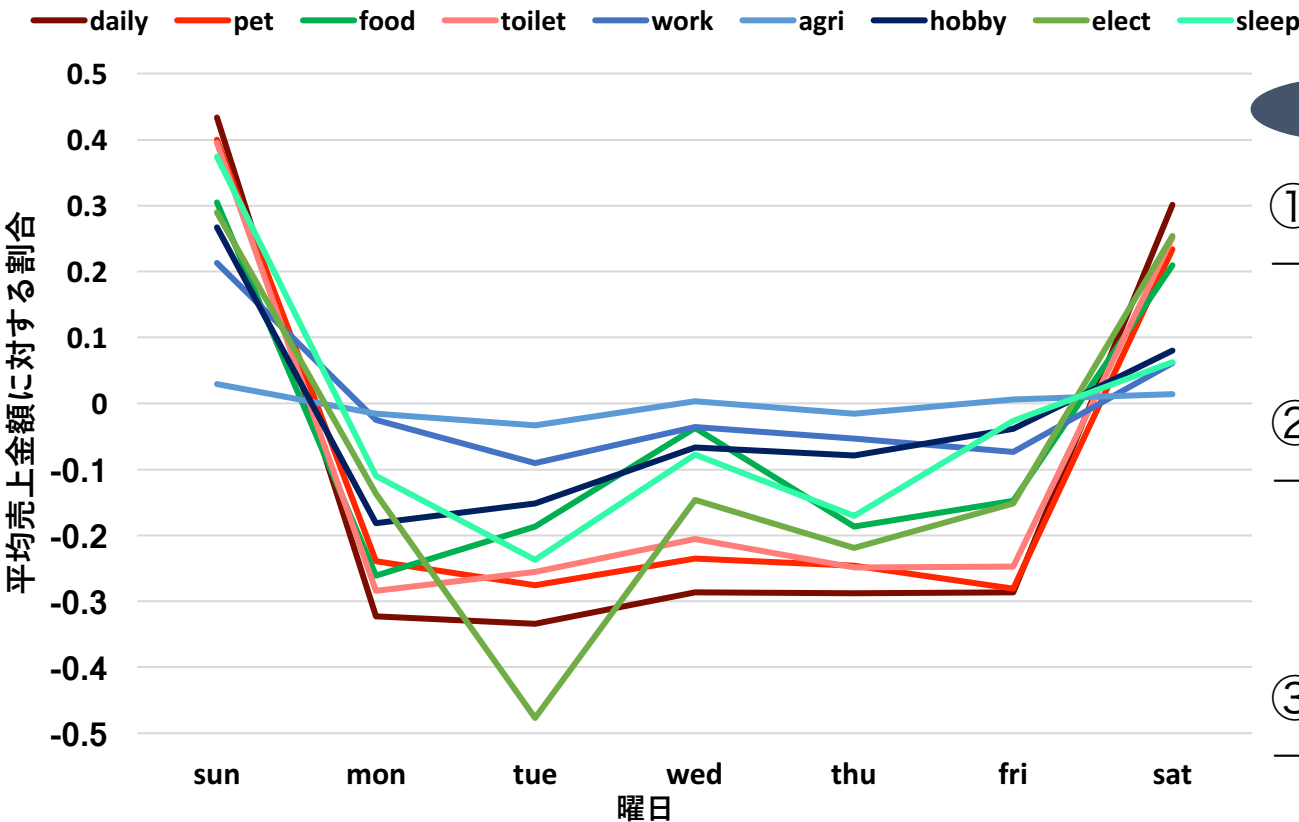
$$y_n = H_n x_n + \varepsilon_n$$

で表現することができる

- 状態推定及び各パラメータの推定にはソフトウェアStanによるハミルトニアンモンテカルロ法を用いた

分析結果 I (曜日成分)

- 曜日成分の中央値をカテゴリ毎に各曜日の平均売上金額で割った値を用いて比較する



カテゴリ毎の曜日効果

傾向

- ① daily, pet, toilet
→ 平日に売上が落ち込み、土日に上がる典型的な傾向
- ② food, elect, sleep
→ 土日に売上が上がり平日は落ち込むが、水曜に一回上がる傾向
- ③ work, agri, hobby
→ 日曜から月曜に売上が落ち込んでからほぼ横ばい傾向

分析結果 II (祝日、祝前日効果)

各類似度の中央値による比較

類似度	daily	pet	food	toilet	work	agri	hobby	elect	sleep
s_{sun}	0.72636	0.60537	0.50023	0.99782	0.99998	0.99593	0.99980	0.99998	0.99984
s_{fri}	0.00048	0.00133	0.00774	0.99819	0.99952	0.99677	0.99777	0.00234	0.99546
s_{sat}	0.03336	0.00413	0.00144	0.12607	0.00088	0.99686	0.99781	0.00079	0.99788

■ 祝日効果

daily, pet, food

→ { 相対的に類似度は低く、祝日が急激に売上に影響しない

toilet, work, agri,
hobby, elect, sleep

→ { ほぼ1に近いいため、直近の日曜の売上で表すことができ、
前述の曜日成分の結果から普段の平日より売上が伸びる

■ 祝前日効果

daily, pet, food, elect

→ { 祝前日効果はほぼないと言える

toilet, work

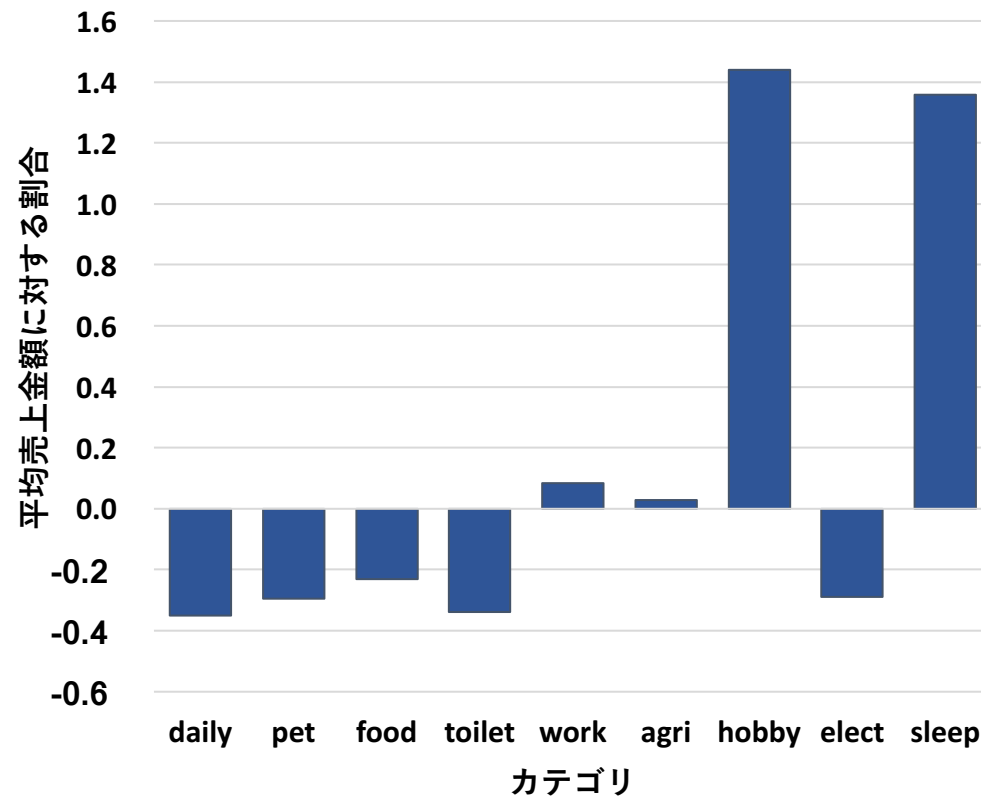
→ { 金曜との類似度が高く、土曜との類似度が低いことから
金曜の売上で表すことができ、前述の曜日成分の符号から
普段の平日より売上は落ちる

agri, hobby, sleep

→ { 両日に対して類似度が高く、前述の各曜日成分の符号と
絶対値の大小から普段の平日より売上が上がる

分析結果Ⅲ (天候成分)

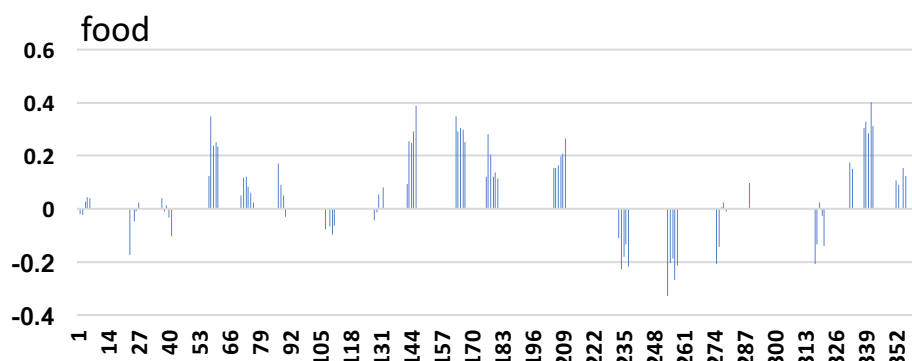
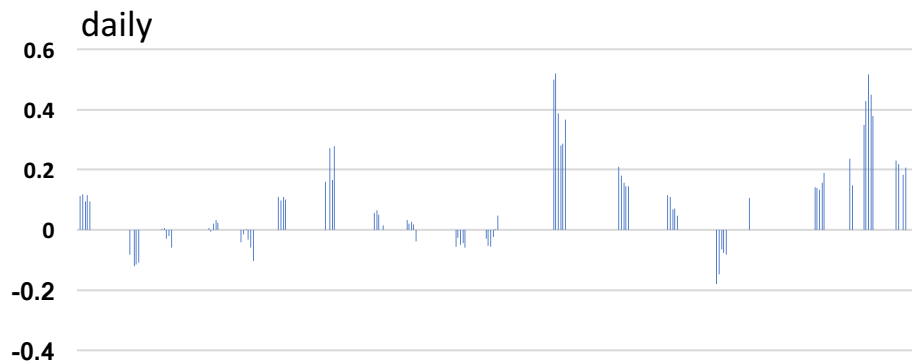
- *c_rain*の中央値をカテゴリ毎に雨天の日の平均売上金額で割った値を用いて比較する



カテゴリ毎の天候効果

- daily, foodのような日常的に売れ筋の良いようなカテゴリは天候が優れない日に売上が下がる傾向にある
- 一方で、hobby, sleepのような普段頻繁に買わないようなカテゴリは悪天候により、売上が上がることを示している
- work, agriなどは、その売上があまり天候に影響されることがないことを示している

分析結果Ⅳ（イベント成分）



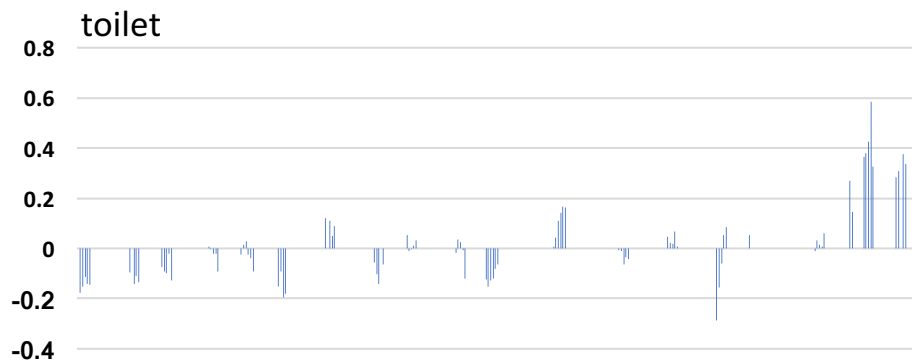
イベント開催日の c_event_t の符号別割合

カテゴリ	効果（日数）	
	正	負
daily	63	28
pet	40	51
food	59	32

- デシルランク 1 に属し日常的に売上の高い3カテゴリだが、petのみ負の効果の日数が正の効果の日数を上回る結果となった
- dailyとfoodの2カテゴリは、正の効果を示す日数の方が多いが、その値が売上金額の半分以上（0.5以上）を占める日数はさほど多くないことがわかる

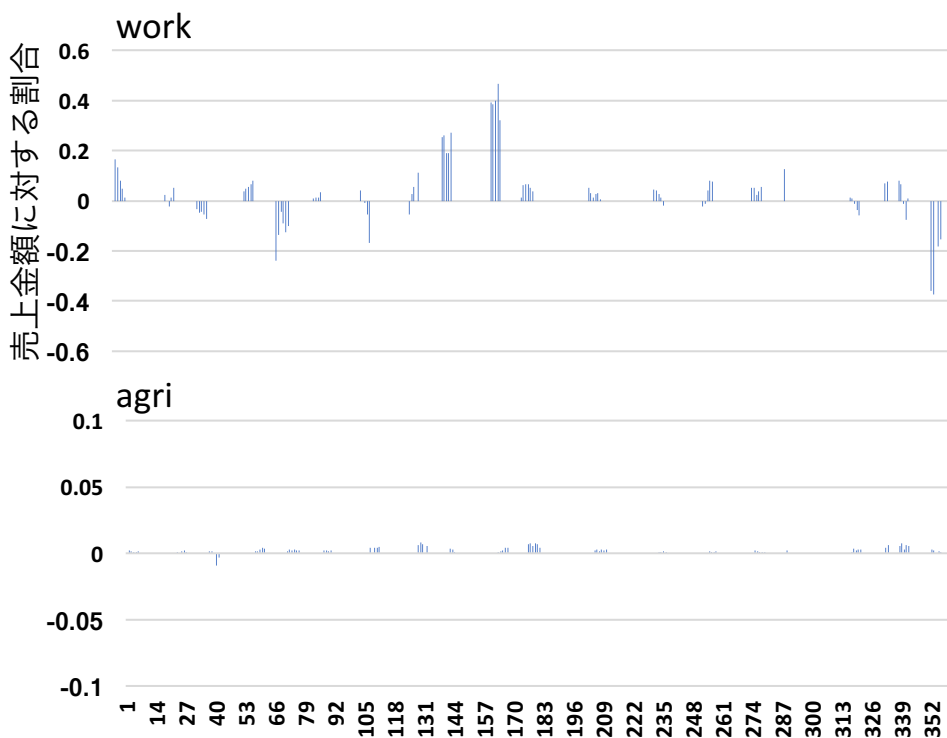
イベント開催日の売上に対する c_event_t の中央値の割合

分析結果Ⅳ（イベント成分）



イベント開催日の c_event_t の符号別割合

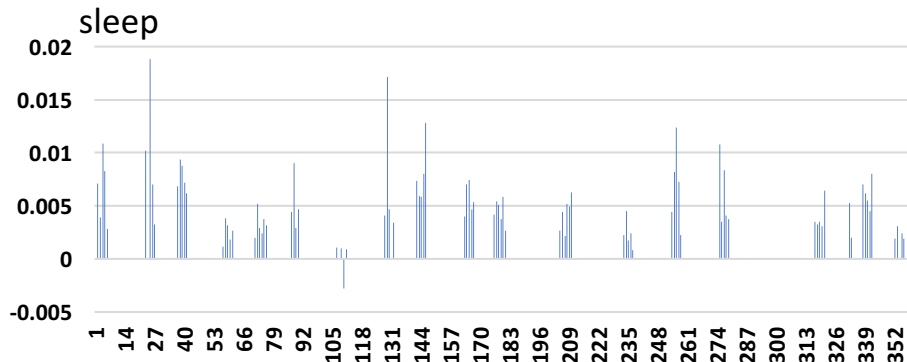
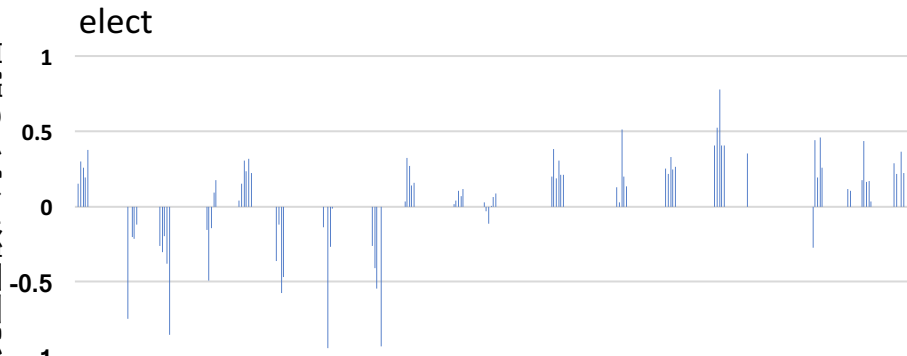
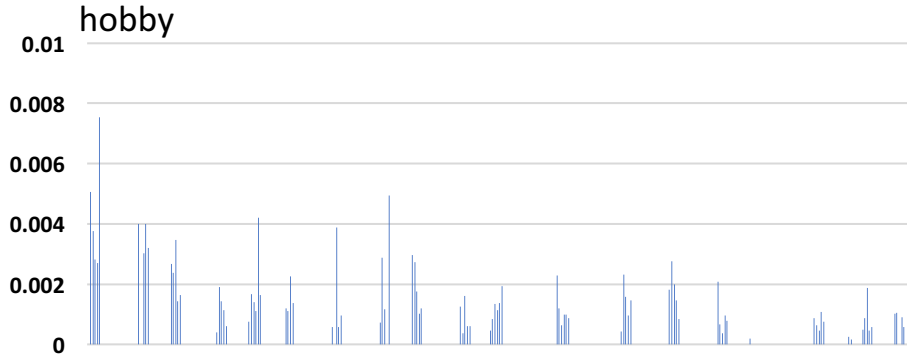
カテゴリ	効果（日数）	
	正	負
toilet	41	50
work	63	28
agri	84	7



- 同じデシルランクに属する3カテゴリだが、イベント開催による効果は三者三様である
- toiletとworkは対照的な効果の現れ方であり、agriに関しては、イベント効果はほぼ存在していないことがわかる

イベント開催日の売上に対する c_event_t の中央値の割合

分析結果Ⅳ（イベント成分）



イベント開催日の c_event_t の符号別割合

カテゴリ	効果（日数）	
	正	負
hobby	91	0
elect	64	27
sleep	89	2

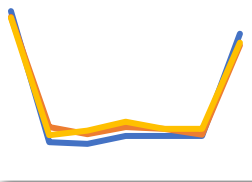
- これらの3カテゴリは、全てイベント効果が正の符号を示す日数が多くなる傾向があることがわかる
- しかし、hobbyとsleepの売上金額に対するイベント効果の割合はかなり小さい値となっている

イベント開催日の売上に対する c_event_t の中央値の割合

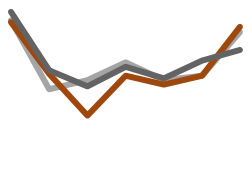
考察 I

■ カテゴリ毎の曜日による売上の変動（分析結果 I）

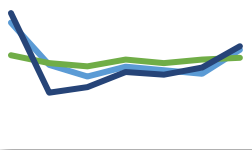
傾向①：小売店としては典型的な売上パターンとなっており該当するカテゴリ(daily, pet, toilet)も日常的に売れているカテゴリであることから、背景知識と一致すると言える。



傾向②：土日に売上が上がり平日は落ち込むという傾向は傾向①と類似する点であるが、水曜日に一度売上が上がる（他の平日と相対的に）という特徴がある。一概に判断することはできないが、水曜日に店舗独自のセールなどの実施していることが想定される。



傾向③：日曜日から月曜日に売上が下がって以降ほぼ横ばいの状態が続く売上パターンとなっている。該当するカテゴリ(work, agri, hobby)と合わせて考えると、購買する顧客層が限定されているかつ、頻繁に購買がされないカテゴリであることがこのような傾向になった要因であると考えられる。



考察 II

■ 祝日・祝前日による売上への影響（分析結果 II）

祝日効果

- ◆ デシルランクが上位のカテゴリ (daily, pet, food) は、祝日効果が正に効いているが、他カテゴリに比べるとその値は高くない。このことは基本的な売上金額が一週間を通して高いために祝日による急激な売上増加に繋がっていないことが要因であると考えられる。

祝前日効果

- ◆ 祝前日効果があるカテゴリのうち、売上が上がるカテゴリ (agri, hobby, sleep) と下がるカテゴリ (toilet, work) に分かれたことについて、該当するカテゴリが顧客が頻繁に購買する商品かどうかといった要因が反映されていると考える。

考察 III

■ 天候の善し悪しによる売上への影響の差

効果あり（負）：該当するカテゴリ (daily, pet, food, toilet, elect) は普段からの売上が高いカテゴリであり、天候不順によって売上が悪くなる傾向は小売業の背景知識とも一致すると言える。



効果あり（正）：上記の考察とは対照的に普段頻繁に購買されないカテゴリ (hobby, sleep) であり、対象店舗がショッピングモールに内在する店舗であることを踏まえると、車での来客が増え非計画購買が発生しやすいカテゴリであると推測される。



効果なし：該当カテゴリ (work, agri) は、購買する顧客層がある程度限定されたカテゴリであると思われるため、天候に影響されにくい結果となったと考えられる。



考察IV

■ イベント開催による売上への影響の差

正の効果：正の効果ではあるが、その値が小さくほとんど効果がないようなものも見受けられたことについて、それらのカテゴリ(work, agri, hobby, sleep)は頻繁に購買されないようなカテゴリである。これまでの考察も合わせて考えると、これらのカテゴリは外部的な要因による影響を受けにくいカテゴリであると考えられる。

負の効果：対象の店舗のヒアリングから、近隣イベントの開催は来店者数に関係する知見が得られていることから、負の効果が多くなることはこうした主観的知見と一致しない。本研究では、イベントの変数として0or1の値が入力として与えられており、適切でなかったと考えられる。

まとめと今後の課題

まとめ

- 本研究では、状態空間モデルによって、ホームセンターにおけるカテゴリ毎の売上データを店舗の外部要因を加味した複数の要因に分解した
- 上記から、同一店舗でもカテゴリによる売上の要因の差を見出すことができた。これにより、実際の店舗の経営戦略（在庫発注など）に役立つことが期待される

課題

- 状態空間モデルを構成する各要因の定式化に不十分な点が存在する。特に近隣のイベント効果の入力変数としては、イベント毎に入場者数などの差があることから、その状況を加味した変数作成を行う必要がある。
- また、本研究の対象店舗は1店舗であり、モデルの有用性を提言するためには今後、複数店舗を用いて比較する必要がある。

参考文献

1. 角田孝昭, 吉田光男, 津川翔: “状態空間モデルを用いた検索トレンドとページビューからの自動車販売台数の予測”, 人工知能学会全国大会, Vol.29, pp.1-4 (2015)
2. 山口類, 土屋映子, 樋口知之: “状態空間モデルを用いた飲食店売上の要因分解”, オペレーションズ・リサーチ, Vol.49, pp.316-324 (2004)
3. 本橋永至, 磯崎直樹, 長尾大道, 樋口知之: “状態空間モデルによるインターネット広告のクリック率予測”, オペレーションズ・リサーチ, Vol.57, pp.574-583 (2012)
4. 山根智之, 菅原光太郎, 西村直樹, 小林健, 吉田佑輔, 高野祐一, 中田和秀: “時系列モデルによる商品販促の分析”, オペレーションズ・リサーチ, Vol.61, pp.65-70 (2016)
5. 近藤文代: “デイリーPOSデータにおける曜日変動および値下効果の抽出”, オペレーションズ・リサーチ, Vol.44, pp.167-183 (1999)
6. 松浦健太郎: “Wonderful R2 StanとRでベイズ統計モデリング”, 共立出版 (2016)
7. 赤池弘次, 尾崎統, 北川源四郎: “統計科学選書 5 時系列解析の方法”, 朝倉書店 (2003)