

# ホームセンターのウェブサイトの購買履歴を利用した シーズン商品の分析

申請中 中央大学大学院 \*尾崎玲央奈 OZAKI Reona  
05000907 東海大学 大竹恒平 OTAKE Kohei  
01405390 中央大学 生田目崇 NAMATAME Takashi

## 1. 研究目的

ホームセンターは近年、ドラッグストアなどの成長により集客商材を奪われている。そこで、市場規模が増加し続けている EC 市場への積極参入が必要だと考えられる。本研究では、あるホームセンターの EC サイトを取り上げ、地域、時期による販売の特性を明らかにする。そして、地域ごとの季節商品の売行きを在庫管理や商品 PR への活用を提案する。

## 2. 使用するデータ

本研究では、あるホームセンターから提供いただいたインターネット販売の POS データを用いて、購買分析を行う。また、気象情報に関しては気象庁のウェブサイトに掲載されているデータを利用した [1]。

- 対象期間 2017/4/1~2018/3/31
- 年間売上金額 約 15 億円
- 年間売上件数 約 34 万件
- 商品カテゴリ数 56 カテゴリ

## 3. 分析と考察

はじめに、気温の変化に基づき都道府県のクラスタリングを行い、6つの都道府県クラスタに分類した。次に、各商品の売上の推移についてクラスタリングし、3つの商品クラスタに分類した。次に、都道府県クラスタと商品クラスタを掛け合わせた 18 クラスタで重回帰分析と決定木分析を行い、これらの結果より季節商品が売れる気候条件を考察する。

### 3.1. クラスタ分析

クラスタ分析は互いに似ているデータをクラスタという部分集合に分ける分析手法である。

本研究では 47 都道府県を一週間ごとの平均気温 52 週分を変数とした。これを基に非階層型クラス

タ分析を行い、都道府県を 6 つのクラスタに分類した。各都道府県の所属を表 1 に示す。

表 1: クラスタごとの所属都道府県一覧

番号	所属都道府県
1	沖縄県
2	愛媛県, 佐賀県, 大分県, 和歌山県, 宮崎県, 鹿児島県, 長崎県, 熊本県, 静岡県, 福岡県, 高知県, 大阪府, 兵庫県
3	愛知県, 広島県, 神奈川県, 奈良県, 京都府, 香川県, 千葉県, 岡山県, 三重県, 東京都, 岐阜県, 山口県, 徳島県
4	岩手県, 北海道, 青森県
5	茨城県, 群馬県, 埼玉県, 山梨県, 滋賀県, 石川県, 鳥取県, 島根県, 栃木県, 富山県, 福井県
6	宮城県, 山形県, 秋田県, 新潟県, 長野県, 福島県

商品の週ごとの売上の比率に関しても同様に、非階層型クラスタリングを行い、商品を「夏商品」、「冬商品」、「通年商品」の 3 つのクラスタに分類した。

### 3.2. 重回帰分析

重回帰分析は 1 つの目的変数を複数の説明変数で予測する多変量解析の手法の 1 つである。目的変数  $y$  は以下の式で示される。

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i \quad (1)$$

ここで、 $y$  は予測したい売上金額 (目的変数)、 $x_i$  は売上金額に影響を与える要因 (説明変数)、 $\beta_i$  は各説明変数に対するパラメータ ( $\beta_0$  は切片) を示す。また、分析を行う際、AIC 基準 (赤池情報量基準) による変数減少法を用いて、説明変数を選択した。

本研究では、目的変数を商品が購入された都道府県ごとの同一商品の一週間の売上金額の合計。説明変数は、気象に関するデータから「気温」、「湿度」、「降水量」、「降雪量」、「風速」、「蒸気圧」、「日照時間」、「一週間前との気温差」、「不快指数」、「体

感温度」，さらに商品カテゴリごとに切片を変化させるために「商品カテゴリ」を使用した。

- 「一週間前との気温差」—その週の平均気温から前の週の平均気温を引いた値。
- 「不快指数」—夏の蒸し暑さを数量的に表した指数で，以下の式で表される。

$$DI = 0.81T + 0.01H \times (0.99T - 14.3) + 46.3 \quad (2)$$

ここで， $DI$  は不快指数， $T$  は乾球気温 ( $^{\circ}\text{C}$ )， $H$  は湿度 (%) である。

- 「体感温度」—人間の肌を感じる温度の感覚を，定量的に表したもので，以下の式で表される。

$$L = T - 4 \times \sqrt{v} \quad (3)$$

ここで， $L$  と  $T$  はそれぞれ体感温度と乾球気温 ( $^{\circ}\text{C}$ )， $v$  は風速 ( $\text{m/s}$ ) である。

「都道府県クラスタ5」の「夏商品」の重回帰分析を行った結果，表2に示す変数が選択された。

表 2: 重回帰分析による結果

説明変数名	偏回帰係数
気温	0.638
風速	0.082
蒸気圧	0.222
日照時間	0.071
湿度	0.085
不快指数	-0.814

「都道府県クラスタ5」の「夏商品」に関しては，6つの変数が選択され，どの変数も5%以下で有意となった。結果より，初夏や春の比較的に気温が高い日に売上が伸びると思われる。これは，暖かくなっていくタイミングで暑さ対策の道具や夏に向けたレジャー用品などが購入されるためだと推測される。

### 3.3. 決定木分析

決定木分析とは定められた目的変数に対して識別する説明変数とその基準を求める分析手法であ

る。結果を木構造で表示するため決定木分析と呼ばれる。

目的変数と説明変数は，基本的には重回帰分析と同様としたが，夏商品の分析の際には冬の購買データを除いて分析を行った。「都道府県クラスタ5」の「夏商品」の売上高に関する回帰木を図1に示す。分析結果より週の平均気温が  $14.1^{\circ}\text{C}$  以上になると売上が上がるが，そのなかで平均風速が  $4.394\text{m/s}$  以上の週は売上が少ない。しかし，対象の地域で一週間の平均風速が  $4.394\text{m/s}$  以上となった週はほとんどないことがわかった。したがって，都道府県クラスタ5の地域では，夏商品は  $14.1^{\circ}\text{C}$  を境に売れるようになり，条件を満たす4月頃から夏商品の在庫を増やす，販売促進を行うことが適切だと考えられる。

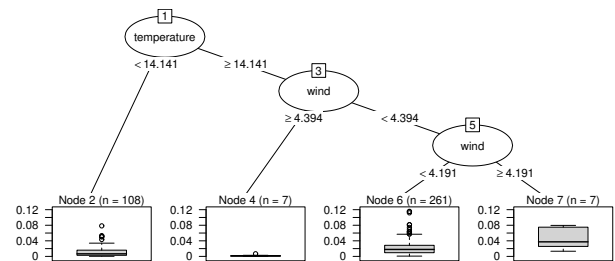


図 1: 都道府県クラスタ5の夏商品の決定木分析の結果

## 4. まとめ及び今後の課題

本研究では季節に注目して地域と商品のクラスタリングを行い，そのクラスタごとの売上の変動の要因を分析した。分析によって，ECショップにおいても，気象情報が売上に変動を与える要因になることが明らかになった。今後の課題として，商品のより細かい分類と，より詳細な売れ始める条件を改善する余地がある。

### 謝辞

本実験に使用したデータの提供に対して，某ホームセンター様に深く感謝致します。また，本研究はJSPS 科研費 19K01945 および 17K13809 の助成を受けたものです。

### 参考文献

[1] 気象庁 過去の気象データ・ダウンロード  
<https://www.data.jma.go.jp/gmd/risk/obsdl/>