

# 時系列売上推定データを用いた食料品店の閉店確率予測モデル

関口達也・貞広幸雄

## Model for prediction the probability of groceries with temporal estimating sales data

Tatsuya SEKIGUCHI and Yukio SADAHIRO

**Abstract** In these days, the number of people that have difficulty in shopping groceries is increasing, due to the closing of grocery stores and super market around them. This study aims to predict the probability of each stores' closing. First, we estimate the sales of each store with the data of population distribution and consumption expenditure for groceries. Second, using that data, we propose the method for prediction the probability of stores by partial logistic model.

**Keywords:** フードデザート (food desert), 閉店予測 (prediction of store closing), 売上推定 (estimation of sales)

### 1. はじめに

「フードデザート」という言葉が注目されている。これは、食料品店の相次ぐ撤退・閉店により店舗への近接性が悪化し、食料品の購入が困難な状況を指す。2000 年以前から海外事例が顕在化し、日本でも買物難民の推定値が 600 万人という国からの発表<sup>1)</sup>や、食料品アクセスマップ<sup>2)</sup>の作成などをうけ、社会的問題としての認識、対策検討が行われつつある。

フードデザートの主な要因の一つに、食料品店の閉店がある。アクセス可能な地域に食料品店が存在しなければ、住民は配達サービスや通信販売、公共交通利用など、より費用の掛かる購買手段を選択する必要がある、経済的・社会的に弱い立場の住民の中には、その手段を全て失ってしまう可能性があるものもある。この問題に対応するには、店舗の閉店を事前に予測し、適切な施策を講ずる必要がある。

店舗の閉店予測に関する既存研究には、まず、西

岡ら (2010)<sup>3)</sup>があり、大規模店舗商業施設の立地シミュレーションを行う中で、生存曲線による店舗の存続年数に応じた生存率を明らかにした。しかし、両者の関係の傾向把握に留まっており、個々の店舗の閉店確率予測という点には十分に応えていない。

また、崔 (2012)<sup>4)</sup>は、1000 m<sup>2</sup>以上の食料品店について、ロジットモデルにより、店舗の立地、周辺競合店舗の立地動向などの地理的要因から個々の店舗の閉店確率推定を可能にした。しかし、モデル自体が静的なものであり、分析期間中の店舗の閉店時期や変数の時系列的な変化を考慮に入れていないことから、将来予測への適用は困難である。以上を踏まえ、本研究では、店舗撤退の意思決定要因の一つである個々の店舗売上を、店舗規模と周辺の人口分布変遷から時系列的に推定し、近い将来を含めた閉店確率を、時系列的に予測する手法を提案する。その結果を、問題が発生しうる地域での発見や、問題の発生可能性が高い地域における事前対応策の検討に役立てていく事を目的とする。

---

関口：〒113-8656 東京都文京区本郷 7-3-1

東京大学大学院 工学系研究科 都市工学専攻

Phone: 03-5841-6259

E-mail: ta-sekiguchi@ua.t.u-tokyo.ac.jp

## 2. 本研究の流れと利用データ

分析の流れとして、まず、近年の食料品店の閉店状況の概観を行う。そして、修正ハフモデルにより平成 12・17・22 年の各地域から各店舗への買物出向人口を計算、それに各地域の食料品の消費額を乗じて、店舗の売上を推定する。その後、それを共変量とする閉店確率予測手法の説明を行う。

分析対象は、関東地方で食料品店への近接性の悪い居住者が多い<sup>(注 1)</sup>、群馬県・栃木県・茨城県の 3 つとした。店舗データは、株式会社商業界より毎年一回発刊の、日本スーパー名鑑（2001~2011 年版）により、2000~2010 年度に存在する各食料品店の売場面積、住所、閉店の有無、既に閉店した場合は閉店年を集計したものである。なお、本研究での「食料品店」の定義は、上記の名鑑で扱われる、食品主体<sup>(注 2)</sup>のスーパー、GMS（7000 m<sup>2</sup>以上）、生協・農協、ディスカунストアである。また、店舗位置は、住所を基に号レベルで緯度経度情報を取得した。人口データについては、平成 12・17・22 年国勢調査の町丁・字等別集計結果を用いて、町丁目コードや名称から各年度のマッチングを行っている。

## 3. 対象地域内の食料品店閉店状況の概観

図 1 に、2000~2010 年の食料品店数の推移を、図 2 に 2010 年に存在する食料品店と、そのうち 2000 年以降に閉店したもの、また 2000~2005 年、2006~2010 年の期間に閉店した食料品店の分布を示す。各県の店舗数の推移は、それぞれやや異なり、群馬県では経年的に減少傾向がある一方で、他 2 県では、2003 年以降増加傾向にある。

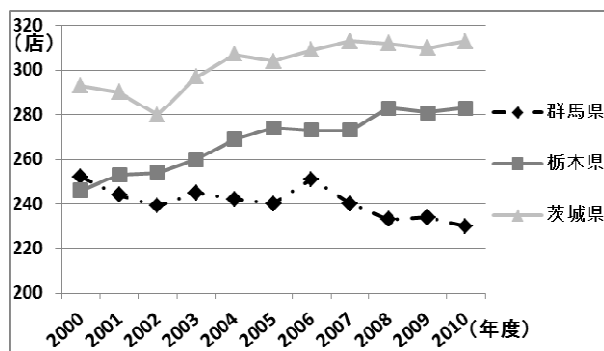


図 1 対象地域の食料品店数の推移

閉店店舗は、2010 年時点で多く店舗の立地する前橋、高崎、宇都宮、水戸、日立など各県の中核的都市に多くみられる。これらの地域では、2000 年以降の食料品店の開店も多く、店舗の新陳代謝が盛んである事がわかる。しかし、その一方で、いずれの県にも店舗が密集せずまばらに立地し、2000 年以降に開店した店舗が少ないにも関わらず、食料品店の閉店がみられる地域がある。このような地域では、一つの店舗の閉店が周辺住民の食料品店へのアクセス悪化につながると考えられる。その閉店の今後の有無を予想する事は、今後の閉店可能性が高い場合、食料品の他の経路からの入手可能性の調査など、事前の対応策を検討するためにも重要であろう。

## 4. 閉店確率予測モデルの手法について

### 4-1 修正ハフモデルを用いた店舗の売上推定

まず、モデルの共変量である各店舗の売上推定を行う。修正ハフモデルにより、各町丁字から 40 km<sup>(注 3)</sup> 以内にある各店舗への出向確率を求めた。

$$P_{ij} = \frac{A_j/d_{ij}^2}{\sum_{j=1}^n A_j/d_{ij}^2} \quad \begin{array}{l} P_{ij} \cdots \text{町丁字 } i \text{ から食料品店 } j \\ \text{への出向比率} \\ A_j \cdots \text{食料品店 } j \text{ の面積} \\ d_{ij} \cdots \text{町丁字 } i \text{ から食料品店 } j \text{ の距離} \end{array} \quad (1)$$

(1)式により算出した、各店舗への出向割合に応じた町丁字の人口を乗じることで、各町丁目から各店

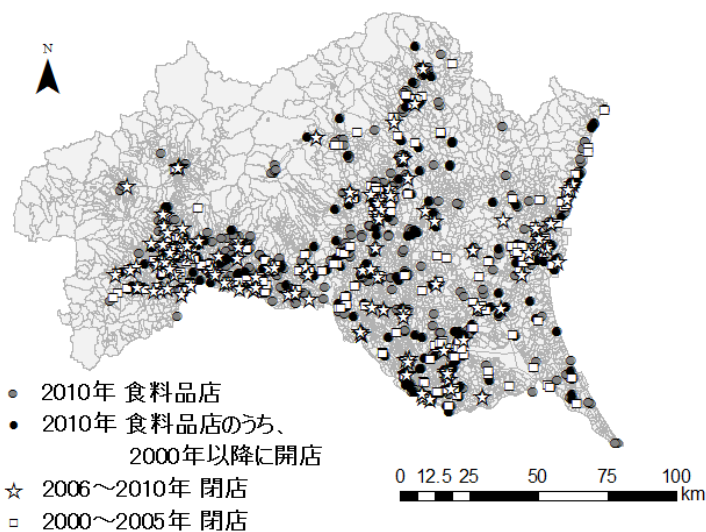


図 2 対象地域の食料品店の開閉店と 2010 年の分布

舗への買物出向人口を求めた。さらに、各町丁目の平成 12, 17, 22 年の買物出向人口に対応させ、各県の平成 11, 16, 21 年の全国消費実態調査による食料品<sup>(注 4)</sup>の消費支出額を乗じて計算した、各町丁目から各店舗への食料品支出額を店舗ごとに集計し、その合計値を各店舗の各年度の売上推定値とした。

#### 4-2 部分ロジスティックモデルによるモデル化

前項で推定を行った各店舗の売上推定値を共変量として、閉店確率予測のモデルを構築する。

部分ロジスティックモデルは、Efron (1988)<sup>5)</sup>や、辻谷ら (2005)<sup>6)</sup>による研究が行われており、時間により変化する共変量を扱い生存データの解析を行う上で有効な手法として着目されている。

店舗  $i$  のある観測時点  $t_{il}$  について、 $l$  番目の時間区間に対する中央値を  $y_{il} = (t_{i(l-1)} + t_{il})/2$  とする。観測起点を  $t_{i0}=0$  とし、2000 年以前から存在した店舗については、2000, 2005, 2010 年に観測を行うものとする。途中の期間で店舗が閉店した場合には、以降の観測は行われない。また、区間途中で出店した店舗については、そこを観測起点とし、以降閉店するまで、2000 年以前からの店舗と同様に観察を行う。つまり、店舗の開閉店の時期により時間区間  $l$  の数は異なり、ここで、店舗  $i$  の最後の時間区間を  $L_i$  とする。また、店舗  $i$  の各観測区間の共変量として、区間  $l$  の冒頭の売上  $x_{il}$ <sup>(注 5)</sup> と、上記の  $y_{il}$  を用いる。

上記の指標を用いて、 $l$  番目の時間区間のハザード関数を以下の様に定義する。

$$h_{il} = 1/1 + \exp\{-(\gamma a_{il} + \beta_0 + \beta_1 x_{il})\} \quad (2)$$

( $l=1, 2, \dots, L_i$ )

$\gamma, \beta_0, \beta_1$  は、時間区間によらず共通の各共変量に対するパラメータである。

店舗  $i$  の  $l$  番目の時間区間での存続性指標を、

$$\delta_{il} = \begin{cases} 1 & (\text{店舗} i \text{ が時間区間} l \text{ で閉店}) \\ 0 & (\text{その他}) \end{cases}$$

$$\delta'_{il} = \begin{cases} 1 & (\text{店舗} i \text{ が時間区間} l \text{ で観測打ち切り}) \\ 0 & (\text{その他}) \end{cases}$$

と定義し、ここまでの各時間区間と観測時点、共変

量との対応関係を、図 3 に示す。

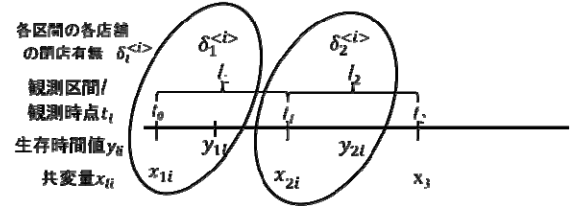


図3 共変量と時間区間、存続性指標の関係性

また、店舗  $i$  の時間区間  $l-1$  までの、死亡または観測打ち切りに関する指標  $A_{il}$  を以下の様に定義する。

$$A_{il} = (\delta_{i1}, \delta'_{i1}, \delta_{i2}, \delta'_{i2}, \dots, \delta_{i(l-1)}, \delta'_{i(l-1)})$$

時間区間  $l$  の冒頭まで閉店のなかった店舗は、時間区間  $l-1$  までは閉店がなく、観測も打ち切られないため、 $A_{il}=(0,0,0,0,\dots,0,0)$  となる。 $A_{il}$  のもとで、次の時間区間  $l$  における  $\delta_{il}$  はベルヌーイ分布  $B_e(1, h_{il})$  に従う事となる。また、 $A_{il}$  を  $\delta_{il}$  まで広げた  $A'_{il}=(0,0,0,0,\dots,0, \delta_{il})$  を用いて、 $\delta'_{il}$  は、推定されるパラメータ  $\gamma, \beta_0, \beta_1$  に依存しない確率分布  $p(\delta'_{il}|A'_{il})$  に従うものとする。すると、ある店舗  $i$  が区間  $L_i$  で閉店が観察される確率分布は、以下の様に表せる。

$$\left\{ \prod_{l=1}^{L_i-1} (1 - h_{il}) \right\} \times (h_{il})^{\delta_{il}} (1 - h_{il})^{1-\delta_{il}} \times \left\{ \prod_{l=1}^{L_i} p(\delta'_{il} | A'_{il}) \right\} \quad (3)$$

これより、全  $n$  店舗に関する対数尤度は、以下になる。

$$\ln L = \ln L(\gamma, \beta_0, \beta_1) + \sum_{i=1}^n \sum_{l=1}^{L_i} \ln p(\delta'_{il} | A'_{il}) \quad (4)$$

ただし、上式  $\ln L(\gamma, \beta_0, \beta_1)$  は、次で表される。

$$\ln L(\gamma, \beta_0, \beta_1) = \sum_{i=1}^n \left\{ \sum_{l=1}^{L_i-1} \ln(1 - h_{il}) + \delta_{il_i}(h_{il_i}) + (1 - \delta_{il_i})(1 - h_{il_i}) \right\} \quad (5)$$

また、時間区間が 1 つしか存在しない店舗についても、同様に、ベルヌーイ分布に従う仮定が成り立つとすると、(5)式における和をとる際には  $L_i=1$  として、右辺の括弧内第 2 項、第 3 項を考えればよい。そして、(5)式を最大化するパラメータ  $\gamma, \beta_0, \beta_1$  の値を推定し、実際の共変量の値により、各店舗の各時間区間におけるハザード関数値算出が可能である。

ここからは、図3を例にとる。本研究の場合、最新の観測時点( $t_3$ )まで生存した店舗について、その最新観測時点で観測された共変量はパラメータ推定には用いられていない。そこで、次の未来の観測時点 $t_4$ を設定する事にする。本来は、 $t_3$ と $t_4$ の間の時間区間の各店舗の閉店の有無の情報があれば、パラメータの更新が行われるが、ここではその情報は入手できない。しかし、時間区間 $l_1$ ,  $l_2$ の情報をを用いて推定したパラメータは期間によらず共通であるため、 $t_4$ を適切に設定する事で、次の $t_3$ ~ $t_4$ の間の時間区間においても、同じパラメータを用いる事が可能であると考えられる。この上で、(1)式に最新観測時点のパラメータを代入し求めた、 $t_3$ ~ $t_4$ の間の時間区間の店舗 $i$ のハザード関数の値を $h_{i3}$ とすると、時間区間 $l_1$ ,  $l_2$ で閉店がなかった店舗が次の $t_3$ ~ $t_4$ の間の閉店をする条件付き確率 $p_{i3}$ は、 (6)

$$p_{i3} = (1 - h_{i1})(1 - h_{i2})h_{i3}$$

と推定することが可能である。数値計算の結果は、学会当日の発表において提示する。

## 5.今後の課題

今後の課題としては、まず、前述の手法において、将来の閉店予測のために、仮定的に $t_4$ を設定する事になるが、それがどこまで先の時点を設定可能であるのかを検証する事が挙げられる。本研究で扱うモデルでは、新しい情報により各共変量の係数パラメータの値が更新されていくことになるので、その影響が無視できなくなる時点はどこからなのか、検証を行う必要がある。

また、モデルの妥当性の検証のため、ブートストラップ法を用いた、モデル全体の有意性と、共変量の有意性を検証した上で、分析期間の細分化、新たな共変量の追加などにより、モデルの精度向上に努めていきたい。

## 注釈

(1) 食料品店までの距離が 500m を超える居住者の割合が 80%を超えるメッシュ面積の多さから

判断した。

- (2) 食料品の売上が売上高の 15%以上であるもの。
- (3) 平成 22 年道路交通センサスより、各対象県の一般道の昼間非混雑時平均旅行速度の最大値から、自動車により 1 時間で到達可能な範囲の目安として用いた。
- (4) 全国消費実態調査の、収支項目の中で、「食料」項目への消費支出の値から、「外食」項目の消費支出の値を引いたもの。
- (5) 2000~2005, 2000~2010 の間の時点で出店が確認され、食料品店について、前者は 2000 年の、後者は 2005 の人口分布から算出した売上推定値を、共変量とする。

**謝辞：**本研究の国勢調査の数値データ（平成 12, 17, 22 年）とシェープファイル（H17 年）、食料品店の号レベルの緯度経度情報の取得には、東京大学空間情報科学センターの共同研究の一環として提供いただいたものを使用した。

## 参考文献

- 1) 経済産業省 (2010), 地域生活インフラを支える流通のあり方研究会報告書, <<http://meti.go.jp/press/20100514004/20100514004-3.pdf>>, 2012 年 8 月
- 2) 農林水産政策研究所 (2012), 食料品アクセスマップ <<http://cse.primaff.affrc.go.jp/katsuyat/>>, 2012 年 7 月
- 3) 西岡直樹, 加藤博和, 戸川卓哉 (2010), 「出店ダイナミズムを組み込んだ大規模商業施設の立地モデル」, 第 41 回土木計画学研究発表会
- 4) 崔唯爛(2012)「大阪大都市圏における大型食料品店の撤退確率」, 2011 年度 GIS-SA 科研 全体報告会
- 5) Efron,B.(1998), “Logostic regression, survival analysis, and Kaplan-Meier curve”, *J.Amer.Stat.Assoc.*, **83**, pp414-425
- 6) 辻谷将明, 左近賢人(2005), 「時間依存型共変量を伴う生存データの解析」, 応用統計学 **34**(1), pp15-29