

# 入力ニューロンの潜在性に着目した 小売店店舗の非継続来店顧客検知モデルの作成

北島 良三, 遠藤 啓太, 上村 龍太郎

## 1. はじめに

本論文は小売店の非継続来店顧客を検知するモデルについて論じたものである。つまり、将来来店が途絶えてしまう可能性のある顧客の検出を目的とする。本研究では非継続来店顧客の検知に加え、どのような顧客の来店が途絶えてしまうのか、消費者行動についても解釈を試みる。モデル作成にはニューラルネットワークを用い、高精度の予測を狙った。

本研究で解析対象となる小売店は日常生活に密着した小売店であり、顧客に継続的に来店してもらうことを期待している。そのため、この小売店では会員顧客を対象に、日頃の購買に基づいて顧客それぞれにクーポンを発券することにより、顧客の継続的な来店を期待する仕組みを導入している。しかし、クーポン発券を実施しても来店が途絶えてしまう顧客は存在する。継続した来店を望む小売店にとって、クーポン発券という来店を期待する受動的な施策だけではなく、どのような顧客の来店が途絶えてしまうのかを知り、その顧客に対して施策を講じる能動的な対応も必要な事柄である。

このようなことから、来店が途絶えてしまう顧客を早期に検出するモデルを作成し、また非継続来店顧客がどのような購買行動をとっている顧客であるのかを知ろうとする本研究の試みは、小売店に対して有益な情報をもたらすものである。

## 2. データ

### 2.1 データ概要

本研究で作成するモデルの入力には、経営科学系研究部会連合協議会主催の平成 26 年度データ解析コンペティションで提供された“全日食チェーン”の POS データを使用する。コンペティションでは 10 店舗分のデータが提供されたが、本研究においては 1 店舗のデータを解析に用いた。本研究ではこのデータを使用して過去 3 カ月間の顧客行動から 2 カ月後の月の来店可能性を予測するモデルを作成する (図 1)。本研究ではある年の 12 月・1 月・2 月のデータを用いて 4 月の継続来店・非継続来店顧客を予測する。3 月をとばして 4 月の来店を予測することの狙いは、モデルによる予測実施後の 3 月の間に該当顧客に対して継続来店を狙った施策を実施できる時間の確保を想定しているからである。運用のイメージとしては 2 月末日までデータを収集し、3 月 1 日の開店時間までに 4 月中に来店があるか途絶えるか予測を行い、予測で非継続来店と判定された顧客に対して 3 月中に施策を講じる流れである。

データは分析対象店舗の顧客 (総数 6,998 人) を対象とし、12 月・1 月・2 月の 3 カ月の範囲でデータを取得した。このデータには 4 月に来店がある顧客が 2,621 人、来店が途絶える顧客が 4,377 人含まれていた。このデータのうち来店が途絶える顧客群から、3 カ月にわたり表 1 に示す全変数の値がゼロの顧客を除いたところ、来店が途絶える顧客は 2,028 件のデータとなった。そこで、継続来店顧客と非継続来店顧客の比率を 50:50 に調整し、最終的に継続来店・非継続来店顧客合わせて 4,056 件の解析用データを作成した。本研究ではこのデータを、モデル作成区間、検証区間、予測区間の 3 区間にそれぞれ 70%、15%、15% の割合で分割して使用した。

予測に用いたデータでは、全変数がゼロであるデー

きたじま りょうぞう  
東海大学大学院総合理工学研究所  
〒 259-1292 神奈川県平塚市北金目 4-1-1  
えんどう けいた  
東海大学理学部数学科 (2014 年度卒業)  
〒 259-1292 神奈川県平塚市北金目 4-1-1  
かみむら りょうたろう  
東海大学情報教育センター・総合理工学研究所  
〒 259-1292 神奈川県平塚市北金目 4-1-1  
受付 15.7.20 採択 15.11.7

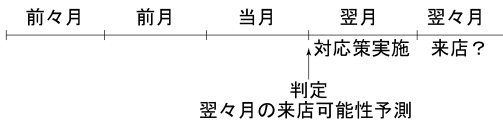


図1 非継続来店顧客検知モデル概要図

タを4月に来店がない顧客群のみから取り除いているため、全変数がゼロのデータが入力された場合“継続来店”という予測結果が出力されてしまうことになるが、これは新規会員に対して“継続来店”の予測出力を狙ってのことである。新規会員は継続して来店しようと思えば会員になったと推測されるので、しばらくの間は来店が継続されると判断した。ただ、この予測出力を得るためのデータ構成により、新規顧客とそうではない顧客の判断が“すべての変数でゼロの値をもつ”というデータだけでは判断ができなくなる。そこで実際の運用においては、すべての変数でゼロの値をもつ顧客に対する予測では、新規会員かそうでないかを調べたうえで、もしも当該顧客が新規会員でない場合は予測結果を反転させて“非継続来店”と最終出力することを想定している。

2.2 変数の作成

非継続来店顧客を検出するために、表1に示す、合計90個から成る変数を用意した。これらは五つの変数ジャンル、3カ月分の変数であり、変数番号1~3は“クーポンの使用率”を、変数番号4~6は“商品の総購入数量”を、変数番号7~48は“POS大分類別の商品購入数量”を、変数番号49~87は“各時間帯における来店回数”を、そして、変数番号88~90は“総来店回数”を示した変数である。POS大分類別の商品購入数量とは、購入した商品の概要を示したものであり、この変数より、どのような商品を購入しているのか知ることが可能となる。また、各時間帯における来店回数は来店した時間についての変数であり、たとえば変数番号49は“時間9(12月)”という変数であるが、これは12月の9時台の来店回数について示した変数である。これらの変数は、クーポン使用率は百分率、それ以外の変数は頻度で値を保持している。なお、POS大分類の14と15が表1に記載されておらず大分類13の次が16になっているが、これはこの二つ(変数は3カ月分なので正確には六つ)の変数が全顧客において値がゼロであり、使用していないためである。

表1 使用した変数

No.	変数名	No.	変数名	No.	変数名
1	クーポン使用率 (12月)	31	POS大分類11 (1月)	61	時間21 (12月)
2	クーポン使用率 (1月)	32	POS大分類12 (1月)	62	時間9 (1月)
3	クーポン使用率 (2月)	33	POS大分類13 (1月)	63	時間10 (1月)
4	総購入数量 (12月)	34	POS大分類16 (1月)	64	時間11 (1月)
5	総購入数量 (1月)	35	POS大分類1 (2月)	65	時間12 (1月)
6	総購入数量 (2月)	36	POS大分類2 (2月)	66	時間13 (1月)
7	POS大分類1 (12月)	37	POS大分類3 (2月)	67	時間14 (1月)
8	POS大分類2 (12月)	38	POS大分類4 (2月)	68	時間15 (1月)
9	POS大分類3 (12月)	39	POS大分類5 (2月)	69	時間16 (1月)
10	POS大分類4 (12月)	40	POS大分類6 (2月)	70	時間17 (1月)
11	POS大分類5 (12月)	41	POS大分類7 (2月)	71	時間18 (1月)
12	POS大分類6 (12月)	42	POS大分類8 (2月)	72	時間19 (1月)
13	POS大分類7 (12月)	43	POS大分類9 (2月)	73	時間20 (1月)
14	POS大分類8 (12月)	44	POS大分類10 (2月)	74	時間21 (1月)
15	POS大分類9 (12月)	45	POS大分類11 (2月)	75	時間9 (2月)
16	POS大分類10 (12月)	46	POS大分類12 (2月)	76	時間10 (2月)
17	POS大分類11 (12月)	47	POS大分類13 (2月)	77	時間11 (2月)
18	POS大分類12 (12月)	48	POS大分類16 (2月)	78	時間12 (2月)
19	POS大分類13 (12月)	49	時間9 (12月)	79	時間13 (2月)
20	POS大分類16 (12月)	50	時間10 (12月)	80	時間14 (2月)
21	POS大分類1 (1月)	51	時間11 (12月)	81	時間15 (2月)
22	POS大分類2 (1月)	52	時間12 (12月)	82	時間16 (2月)
23	POS大分類3 (1月)	53	時間13 (12月)	83	時間17 (2月)
24	POS大分類4 (1月)	54	時間14 (12月)	84	時間18 (2月)
25	POS大分類5 (1月)	55	時間15 (12月)	85	時間19 (2月)
26	POS大分類6 (1月)	56	時間16 (12月)	86	時間20 (2月)
27	POS大分類7 (1月)	57	時間17 (12月)	87	時間21 (2月)
28	POS大分類8 (1月)	58	時間18 (12月)	88	総来店回数 (12月)
29	POS大分類9 (1月)	59	時間19 (12月)	89	総来店回数 (1月)
30	POS大分類10 (1月)	60	時間20 (12月)	90	総来店回数 (2月)

3. 解析手法

3.1 ニューラルネットワークによる予測

本研究では非継続来店顧客の検知モデル作成にニューラルネットワークを使用する。これは高精度の予測結果を得ることを狙ってのことである。ニューラルネットワークは脳の情報処理を模倣したもので、脳細胞であるニューロンのモデルをネットワーク状にした構造をしている。このニューラルネットワークは複雑なデータの処理を得意とするモデルであり、本研究で解析対象とする消費者行動という複雑なデータの解析において効果を発揮するものと期待される。しかし、ニューラルネットワークを用いたモデルには“ブラックボックス”と称される問題が存在する。これは、ニューラルネットワークを用いたモデルはデータを入力すれば予測結果が得られるものの、どのような処理を経てその結果が出力されたのかがわかりにくいという問題を表したものである。高精度の予測のみを目的とする場合、このブラックボックス問題は問題にはならないが、本研究で消費者行動の解釈を試みるように、“どうして”を知る必要がある場合は、これは大きな問題である。

この問題に対応するため、新しいニューラルネットワークを作成し解析に使用した。この新しいニューラルネットワークは“ポテンシャル学習”と名付けられ“Selective Potentiality Maximization” [1] と“Self-Organizing Selective Potentiality Learning” [2] を基に作成した。ポテンシャル学習は、さまざまな状況

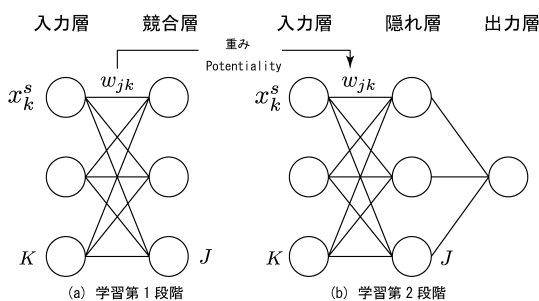


図2 ポテンシャル学習概要図

においても重要な役割を果たす入力ニューロンを意味する“potentiality”の高い入力ニューロン（定義は後記する）に着目して学習を行うニューラルネットワークである。Potentialityの高い入力ニューロンとは、入力データと密接に関係しデータ処理過程に深く関わっているニューロンという意味であり、学習完了後にこのpotentialityの高い入力ニューロンについて調べることで消費者行動に迫ることが可能となる。

### 3.2 ポテンシャル学習

ポテンシャル学習の概要図を図2に示す。図に(a), (b)で示す二つの学習段階が示されているように、ポテンシャル学習は二つの学習段階から成り立っている。一つ目の段階は、データが入力される“入力層”と、データの分類を行う“競合層”から成り立っており、potentialityの高いニューロンを探し出すとともに入力データについての知識を獲得し、入力データを分類する。この段階の学習は自己組織化マップ (self-organizing maps, SOM) が基になっており、学習は教師なし学習で行われる。二つ目の段階は、“入力層”とデータの予測値を出力する“出力層”，そして出力層と入力層を繋ぐ“隠れ層”の3層から成り立っており、誤差逆伝播法 (backpropagation, BP) が基になった学習である。この学習段階ではターゲットを予測する機能を有しており、この段階での学習は教師あり学習で行われる。本研究でのターゲットは4月に顧客が来店しているか否かである。またここでモデルと使用データの関係について補足をを行う。2.1節にて使用データを3分割して使用することについて触れたが、学習第1段階では全入力データが学習され、そして学習第2段階において、モデル作成区間と検証区間のデータを用いて学習が行われる。

学習第1段階で計算される potentiality は、入力ニューロンと競合層のニューロンの重みの分散を基に算出される。図2に示すように  $w_{jk}$  を  $k$  番目の入力ニューロンから  $j$  番目の競合層のニューロンへの重み

とし、 $J$  を競合層のニューロンの個数としたとき、 $k$  番目の入力ニューロンの分散  $V_k$  を

$$V_k = \frac{1}{J-1} \sum_{j=1}^J (w_{jk} - w_k)^2,$$

で定義する。ここで  $w_k$  は  $k$  番目の入力ニューロンからの重みの平均を意味し

$$w_k = \frac{1}{J} \sum_{j=1}^J w_{jk},$$

から計算される。 $k$  番目の入力ニューロンの potentiality を意味する  $\phi_k^r$  は、

$$\phi_k^r = \left( \frac{V_k}{\max_{k=1, \dots, K} V_k} \right)^r,$$

より求められる。分母は、 $k$  が1から入力ニューロンの個数である  $K$  までのときの分散  $V_k$  の最大値である。また、 $r$  は係数であり、potentiality を強調する役割をもっている。よって、 $\phi_k^r$  は、パラメータ  $r$  のときの  $k$  番目の入力ニューロンの potentiality を意味している。本研究で  $r$  は  $0.1 \leq r \leq 1.0$  の範囲の値とした。

このようにして求められた potentiality は本研究では SOM に組み込まれ、第1段階の学習が行われる。SOM の学習では、 $\|\cdot\|$  を距離、 $\mathbf{x}^s$  を  $s$  番目のパターンの入力ベクトル、 $\mathbf{w}_j$  を競合層の  $j$  番目のニューロンへ接続されている重みベクトル、 $x_k^s$  を  $s$  パターン目の  $k$  番目の入力とし、

$$\|\mathbf{x}^s - \mathbf{w}_j\| = \sum_{k=1}^K (x_k^s - w_{jk})^2,$$

により入力データと重みの距離が求められるが、ポテンシャル学習では

$$\|\mathbf{x}^s - \mathbf{w}_j\| = \sum_{k=1}^K \phi_k^r (x_k^s - w_{jk})^2,$$

により距離を求める。距離を求めた後の学習方法は通常の SOM と同様である。距離を求める際に  $\phi_k^r$  を用いているが、このようにポテンシャル学習では potentiality が高いニューロンを重要視するように学習が行われ、この点で通常の SOM と異なっている。

以上の流れで第1段階の学習は終了し、続いて第2段階での学習へと移行する。学習第2段階ではターゲットを予測する働きをし、本研究では BP を基に学習を行う。通常の BP との違いは、入力層のニューロンと隠れ層のニューロンを結ぶ重みの初期値に学習第1段

階で獲得した知識を用いることである。通常、BPの重みの初期値にはランダムに与えられた重みを使用されるが、ニューラルネットワークの学習結果は重みの初期値に左右されるため、ニューラルネットワーク以外の手法により事前に得られた値を重みに使用する [3, 4]、乱数の中から最適な重みを探す [5, 6]、ほかのニューラルネットワーク手法により事前に得られた重みを使用する [7]、などさまざまな試みが行われている。ポテンシャル学習では、学習第1段階で得られた重みをそのまま学習第2段階の重みとするのではなく、重みにpotentialityを乗算した値を使用する。このようにポテンシャル学習の学習第2段階での初期重みに学習第1段階で得られた知識を用いることで、学習第1段階での学習結果に基づいてターゲットを予測することが期待できる。学習第1段階で得られた知識を予測に活用することはこれまで小規模なデータで試みられてきたが [2]、本研究で使用する数千件規模のデータでは初めての試みである。

## 4. 結果と考察

### 4.1 継続来店・非継続来店の予測

#### 4.1.1 各モデルによる予測正解率

図3は非継続来店顧客検知モデルの正解率を示したものである。本研究では継続来店顧客を“0”、非継続来店顧客を“1”としてクラス分けを行ったうえで予測を行っており、図3で示している正解率は、ターゲットの実測値と予測値の正解の割合を表している。このため、予測正解率は全データの予測が間違っていた場合は0を、全データの予測が正解していた場合は1となる。

また、ここで示している結果は、10回データサンプルを変えさせながら予測を行った際の予測区間での平

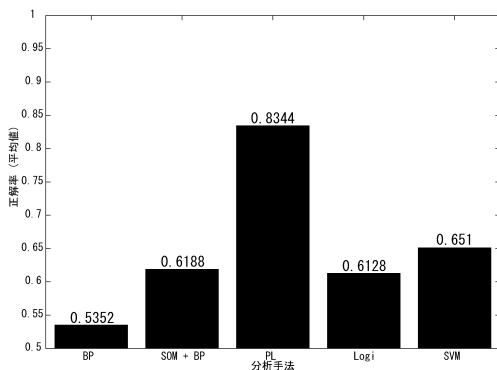


図3 各非継続来店顧客検知モデルの正解率

均値である。サンプルの変更とは、解析に使用した各区間（モデル作成区間、検証区間、予測区間）を各区間の割合はそのままに、使用サンプルをランダムに変化させることである。

予測結果は比較のため、“誤差逆伝播法、BP”、potentialityを使用しないポテンシャル学習と同義である“BPにSOMで得た知識を加えた方法、SOM + BP”、そして“ポテンシャル学習、PL”の三つのニューラルネットワーク手法と、機械学習手法として“サポートベクターマシン、SVM”、そして統計手法として“ロジスティック回帰分析、Logi Reg”を用いて実施した。本研究で比較のための統計的手法としてロジスティック回帰分析を使用したのは、2値のターゲットを予測する場合に一般的に使用される手法であるからである [8]。また、SVMはソフトマージンSVMを使用し、カーネル関数にガウス放射基底関数を用いた。これら各解析手法は数値計算言語のMATLABを使用して実装・予測を行った。

使用した各ニューラルネットワークの構成について、伝達関数を表2に示す。また、各層のニューロンの個数であるが、SOM + BPのSOM部、PLの学習第1段階は出力層に319個ニューロンを使用している。また、BP、SOM + BPのBP部、PLの学習第2段階の隠れ層のニューロンの数は同じく319個、出力層のニューロン数はターゲットの2値を2ビットで表現しているため、2個使用している。319個というニューロンの数であるが、これはデータサンプル数を $n$ として、 $5\sqrt{n}$ より<sup>1</sup>算出した。

図3に示すように、ニューラルネットワーク手法による結果では、BP、SOM + BP、ポテンシャル学習の順で正解率が大きくなり、ポテンシャル学習では $r = 0.3$ の際に正解率が0.8344という値を得た。この順番でニューラルネットワークによる正解率が大きくなるということから、非継続来店顧客の予測には、データから直接予測する(BP)よりも、データから得た知識を使用して予測する(SOM + BP)、そしてさらにpotentialityを導入して予測する(ポテンシャル学習)ことで、高精度の予測ができることが示された。また、

表2 各ニューラルネットワークの構成

伝達関数	SOM + BP		PL
	BP	BP	学習第2段階
隠れ層	双曲線正接関数	双曲線正接関数	双曲線正接関数
出力層	ソフトマックス関数	ソフトマックス関数	線形関数

<sup>1</sup> SOM Toolbox Version2.0 beta [9] のデフォルト設定。

表3 モデル別混合行列

n = 608	予測値										
	BP		SOM + BP		PL		Logi		SVM		
	非継続来店	継続来店	非継続来店	継続来店	非継続来店	継続来店	非継続来店	継続来店	非継続来店	継続来店	
実測値	非継続来店	197.7	106.4	205.5	98.6	<b>273.5</b>	<b>30.6</b>	229.2	70	118.7	180.5
		(0.3252)	(0.1750)	(0.3380)	(0.1622)	(0.4498)	(0.0503)	(0.3770)	(0.1151)	(0.1952)	(0.2969)
	継続来店	176.2	127.7	133.2	170.7	70.1	233.8	165.4	143.4	<b>31.7</b>	<b>277.1</b>
		(0.2898)	(0.2100)	(0.2191)	(0.2808)	(0.1153)	(0.3845)	(0.2720)	(0.2359)	(0.0521)	(0.4558)

表4 各モデルのFPRとFNR

評価尺度	BP	SOM + BP	PL	Logi	SVM
FPR	0.5798	0.4383	0.2307	0.5356	<b>0.1027</b>
FNR	0.3499	0.3242	<b>0.1006</b>	0.2340	0.6033

その他の手法では Logi Reg, SVM の順で正解率が大きくなっていった。

正解率について、作成したモデル全体で確認すると、ポテンシャル学習が最も高い正解率を記録していた。ここで、作成したモデルについて混合行列を用いてさらに詳細に確認する。各モデルの混合行列を表3に示す。

混合行列はモデルがどのような結果を出力しているのかを示したもので、実測値のターゲットが“非継続来店 (ターゲット = 1)”であるデータに対して、予測値が“非継続来店 (ターゲット = 1)”であるパターン (True Positive, TP), 実測値のターゲットが“非継続来店 (ターゲット = 1)”であるデータに対して、予測値が“継続来店 (ターゲット = 0)”であるパターン (False Negative, FN), 実測値のターゲットが“継続来店 (ターゲット = 0)”であるデータに対して、予測値が“非継続来店 (ターゲット = 1)”であるパターン (False Positive, FP), 実測値のターゲットが“継続来店 (ターゲット = 0)”であるデータに対して、予測値が“継続来店 (ターゲット = 0)”であるパターン (True Negative, TN) の4パターンより、モデルの判定特徴を把握できる。表3に示す数値は予測区間のデータ (608 データ) が各パターンの判定となった個数と 608 データに占める割合 (カッコ内の数値) を示しており、また、各モデル手法共に 10 回データサンプルを変更させながら予測を行った結果の平均値である。

表3より、最も非継続来店予測を的中 (TP) させるモデルはポテンシャル学習であること、そして最も来店顧客予測を的中 (TN) させるモデルは SVM であることがわかる。また、予測が外れた場合 (FN と FP) を確認し、モデルの特性について見てみると、SVM 以外のモデルでは継続来店顧客を非継続来店と判定 (FP)

してしまう傾向、SVM では非継続来店顧客を継続来店と判定 (FN) してしまう傾向があることがわかる。

混合行列に加え、継続来店顧客が予測において非継続来店顧客と判定されたかどうかを示す“偽陽性率, False Positive Rate, FPR”と非継続来店顧客が予測において継続来店顧客と判定されたかどうかを示す“偽陰性率, False Negative Rate, FNR”を求めた。これらについて表4に結果を示す。これによると FPR は SVM が最も小さい値であり、FNR は PL が最も小さい値であることがわかった。

“継続来店顧客を非継続来店と判定してしまうこと (FP および FPR)”と“非継続来店顧客を継続来店と判定してしまうこと (FN および FNR)”について考察すると、前者は来店予定顧客に来店を促すことになり、後者は非継続来店顧客に何の施策も行えないということになる。このことより有益なモデルは、表3からは“TP と TN の割合が多く、FN の割合が少ないモデル”, 表4からは“FNR が小さいモデル”, そして、図3に示してある正解率からは“正解率が大きいモデル”, が有益なモデルであるといえ、これらを総合的に判断して本研究では、ポテンシャル学習が顧客の非継続来店可能性予測には最適手法であると判断できた。

#### 4.1.2 ポテンシャル学習の potentiality の活用

ポテンシャル学習による値の予測には学習第1段階で得られた potentiality が活用されている。ここでは potentiality について確認し、モデルがどのような potentiality を活用しているのかを確認する。

図4は  $r = 0.3$  の際にポテンシャル学習の学習第1段階で行われた教師なし学習による学習結果を示している。学習第1段階は SOM が基になっているので、学習結果は入力データの分類結果を意味し、またこれは、データからポテンシャル学習の学習第1段階が獲得した知識を意味している。

図4は学習後の出力層の状態を U-Matrix にて表している。SOM による分類結果は基本的には隣り合うニューロン間は近い距離にあるデータを示している。しかし隣り合うニューロンであっても距離が離れている場合もあるため、U-Matrix はこの距離について色を

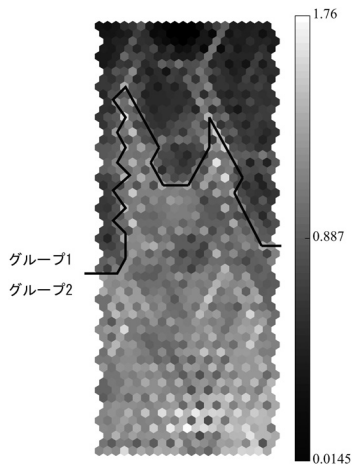


図4 データから獲得した知識、ポテンシャル学習、学習第1段階での学習結果 ( $r = 0.3$ )

用いて奥行情報をもち三次元で結果を示している。白色系の箇所が奥行が高いことを、黒色系の箇所が低いことを示しており、奥行が高い箇所がデータ分類の境界線を意味する。このように図4では色を頼りにデータの境界線を判断し、分類結果を確認することになる。なお、色と距離の対応関係であるが図4右端の数値にそれを示している。

本研究においては色を手掛かりに図中にて実線で示している箇所を境界線と判断し、この境界線より上方を“グループ1”，下方を“グループ2”と命名した。これら各グループについて分類状態を調査したところ、グループ1には1,387人の非継続来店顧客、836人の継続来店顧客、グループ2には641人の非継続来店顧客、1,192人の継続来店顧客が含まれていることがわかった。このことは全非継続来店顧客(2,028人)のうち、約68%がグループ1に、約32%がグループ2に含まれているということであり、グループ1は来店途絶顧客のグループといえる。この来店途絶顧客がグループ1に多く含まれていることについて比率の差の検定を実施して調査したところ  $p < .01$  で有意な差であった。ポテンシャル学習の学習第2段階では学習第1段階で得られた知識を活用することになるので、ここで示されたように、差がある分類結果を学習第1段階で得られたことは、学習第2段階にとって有益となる。各モデルの精度比較の結果、ポテンシャル学習によるモデルが最も高精度であったが、それは学習第1段階で得られたこの知識が活用されているからではないかと判断できる。以上で示されたようにこの節では、ポテンシャル学習が知識を活用してターゲットを予測していることが確認された。

## 4.2 消費者行動

### 4.2.1 ポテンシャル学習を用いた消費者行動の解釈

図5(a)は入力変数のpotentialityを示したものである。図5(a)中の $x$ 軸は変数番号を表したものであり表1の変数番号と一致する。これによると、変数番号6の総購入数(2月)が最もpotentialityが高い、すなわち最重要変数であることが見て取れる。

重要度を変数のジャンルごとに確認すると(表5)、クーポン使用率では変数番号3の2月の使用率が、総購入数量では変数番号6の2月の総購入数量が、総来店回数では変数番号88の12月の総来店回数が、それぞれ上位1位の重要変数であった。総来店回数のみが12月の変数が重要となっているが、potentialityを見てみると変数番号88は0.6602であり、重要度第2位の変数番号90の2月の来店頻度は0.6554と、その差は0.0048であり、ほとんど差がない結果であった。また、クーポン使用率も総購入数量も上位3位までのpotentialityを確認するとあまり差がない結果であり、どの月であっても重要であるといえる。

続いて、継続来店顧客はどのような商品を、どのような時間帯に購入しているのかを、POS大分類と、来店時間から調べていく。図5(b)と(c)は(a)からこの二つの変数ジャンルを抜き出したものである。 $x$ 軸の番号は(a)と一致する。図5(b)より、POSに関する変数は各月で同様のpotentialityをもっており、重要度もほぼ共通しているように見える。各月の重要度上位3位までを確認したところ(表6)、POS大分類では、前々月で変数番号7(大分類1)、11(大分類5)、10(大分類4)、前月で変数番号21(大分類1)、25(大分類5)、24(大分類4)、当月で変数番号35(大分類1)、39(大分類5)、38(大分類4)であり、ともに大分類1(青果)、5(加工食品)、4(惣菜)の順で重要変数が並んでいた。月ごとに重要商品の順位に差がないことから、青果、加工食品、惣菜が重要な商品であることがわかった。

図5(c)に示す来店時間もPOSと同様に各月で似た重要度を示している。各月の上位3位を確認すると(表7)、来店時間では前々月が変数番号56(16時台)、51(11時台)、52(12時台)、前月が65(12時台)、70(17時台)、64(11時台)、当月が77(11時台)、83(17時台)、78(12時台)の順であり、総じて11時台から17時台の範囲で重要な時間帯であることがわかった。この範囲の時間帯は昼食準備から夕食準備の時間帯と解釈することが可能であり、また、この解釈は重要商品が食料品であることから支持される。

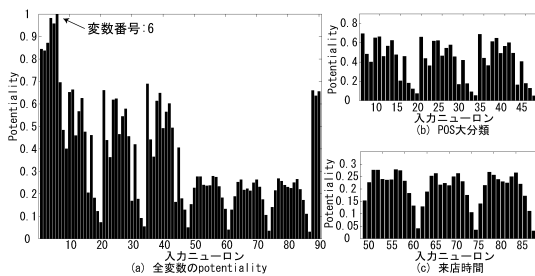


図5 ポテンシャル学習による重要変数

表5 クーポン利用率・総購入数量・総来店回数の重要度 (上位3位: ポテンシャル学習)

順位	クーポン利用率		総購入数量		総来店回数	
	月	Potentiality	月	Potentiality	月	Potentiality
1	2月	0.8711	2月	1.0000	12月	0.6602
2	12月	0.8438	12月	0.9817	2月	0.6554
3	1月	0.8370	1月	0.9576	1月	0.6363

表6 各月の重要POS大分類 (上位3位: ポテンシャル学習)

順位	12月		1月		2月	
	POS大分類	Potentiality	POS大分類	Potentiality	POS大分類	Potentiality
1	大分類1	0.6957	大分類1	0.6604	大分類1	0.6896
2	大分類5	0.6637	大分類5	0.6235	大分類5	0.6484
3	大分類4	0.6524	大分類4	0.6188	大分類4	0.6132

表7 各月の重要来店時間 (上位3位: ポテンシャル学習)

順位	12月		1月		2月	
	来店時間	Potentiality	来店時間	Potentiality	来店時間	Potentiality
1	時間16	0.2788	時間12	0.2633	時間11	0.2683
2	時間11	0.2769	時間17	0.2632	時間17	0.2656
3	時間12	0.2764	時間11	0.2526	時間12	0.2565

得られた重要度のうち、POS大分類と来店時間に關してデータと突き合わせて確認したところ、POS大分類1, 5, 4の3分類合計の購入商品数で全POS大分類の約48% (3カ月分の合計から算出) の購入商品数を占めており、また、来店時間に関しては11時台から17時台の合計来店回数で全時間の約74% (3カ月分の合計から算出) の来店回数を占めていることがわかった。これより、ポテンシャル学習は重要な変数をデータから抽出できていると確認された。また、これらの変数は変数の値が大きいほど、継続来店顧客である傾向があることがわかり、さらに、変数の値から見た継続来店顧客と非継続来店顧客の関係は、五つの変数ジャンル平均で、約77:23 (3カ月分の合計から算出) となっており、おおよそ80:20の関係であることがわかった。

#### 4.2.2 従来手法を用いた消費者行動の解釈

この節では解釈の比較のため、ロジスティック回帰分析にて得られた結果を用いて消費者行動の解釈を試み

る。図6(a)はロジスティック回帰分析で得られた各変数のオッズ比を示している。ここで示しているオッズ比は図3記載の平均正解率を算出するために作成した10個のモデル中で最も正解率が高かったモデル(正解率:0.6283)でのオッズ比である。図中のx軸は変数番号を表したものであり表1の変数番号と一致する。これによると、変数番号74の時間21(1月)が最も影響力が強い、すなわち重要変数であることが見て取れる。

重要度をジャンルごとに確認すると(表8)、クーポン利用率では変数番号3の2月の利用率が、総購入数量では変数番号6の総購入数量(2月)が、総来店回数では変数番号88の総来店回数(12月)が、それぞれ重要な変数であった。総来店回数を除いたほかの変数ジャンルでは重要度第1位・第2位・第3位が2月・1月・12月の順であり、予測月に近い月の購買行動が予測に重要な影響を与えていることを示唆している。

続いて継続来店顧客はどのような商品で、どのような時間帯に購入しているのかを、POS大分類と、来店時間から調べていく。図6(b)と(c)は(a)からこの二つの変数ジャンルを抜き出したものである。x軸の番号は(a)と一致する。これを見ると、図6(b)より、POS大分類では変数番号48の判定当月の大分類16(その他)がほかのPOS大分類の変数よりも大きなオッズ比をもっていることがわかる。各月のPOS大分類の重要度上位第3位までを確認すると(表9)、前々月が変数番号16(大分類10(雑貨))、15(大分類9(菓子))、13(大分類7(乳卵))、前月が変数番号32(大分類12(たばこ))、33(大分類13(米))、30(大分類10(雑貨))、当月では変数番号48(大分類16(その他))、44(大分類10(雑貨))、40(大分類6(パン))であった。各月で共通する商品は雑貨であることがわかるものの、ほかの商品は各月で異なっており、重要商品について傾向を総じて述べるのが難しい結果であった。

また、来店時間について図6(c)より確認すると、最重要変数は変数番号74である1月の時間21であることがわかった。続いて月ごとに上位3位を見ていくと(表10)、前々月が変数番号60(20時台)、61(21時台)、54(14時台)、前月が変数番号74(21時台)、66(13時台)、62(9時台)、当月が変数番号78(12時台)、79(13時台)、75(9時台)の順であった。来店時間の傾向は総じて解釈すると、9時台の朝の時間帯から21時台の夜の時間帯が重要であることがいえる。ただし、これでは全時間帯が重要であるという意

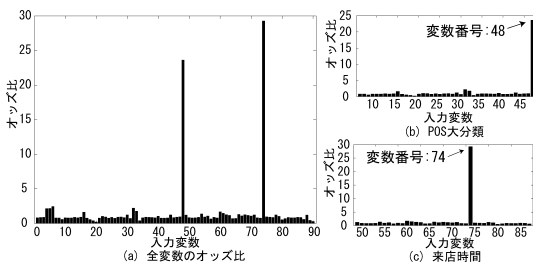


図6 ロジスティック回帰分析による重要変数

味になり、解釈をする意味をもたない結果といえる。

以上のように、各変数の重要度を確認した結果、クーポン使用率と総購入数より予測判定月の近い月の購買行動が重要であること、POS大分類と来店時間より9時台から21時台に雑貨を購入する購買行動が重要であることがわかった。しかし、“なぜ雑貨なのか、なぜこの時間帯なのか”と事柄について解釈することはできなかった。どのような商品を、どのような時間帯に顧客が購入するのかを解釈できなければ、顧客に対する施策を策定することは難しく、このことよりロジスティック回帰分析では消費者行動までは踏み込めない結果であったといえる。

本研究の解析においてロジスティック回帰分析では深い解釈ができない結果となったが、この原因の一つに入力データの問題が考えられる。通常、ロジスティック回帰分析で解析を行う場合、多重共線性問題の発生していないデータが作成・使用される。しかし、本研究において入力データは解析条件を一律にするため、各解析手法共通となっており、またこれは多重共線性について考慮していない。そのため、ロジスティック回帰分析が本来のポテンシャルを発揮できていないと推測することができる。また、本研究では予測においてもロジスティック回帰分析がデータから影響を受けた事柄が散見された。図3の結果を得るために10回サンプルを変更して予測を行ったが、10回中、2回でロジスティック回帰分析の計算が完了しなかった。ロジスティック回帰分析では計算を反復して行いが、この計算が制限回数以内に終了しなかったのである。8回は反復制限回数以内に計算が終了しているのに、この2回の計算で使用されたデータとの相性が悪かったといえる。このようにロジスティック回帰分析は使用するデータを慎重に選ばなければならないという特徴がある。だが、このことは逆にロジスティック回帰分析の弱みともいえ、入力したデータをそのまま扱い、予測・解釈することができる手法に“解析手法の使いやすさ”という点で軍配が上がる。

表8 クーポン使用率・総購入数量・総来店回数の重要度 (上位3位:ロジスティック回帰分析)

順位	クーポン使用率		総購入数量		総来店回数	
	月	オッズ比	月	オッズ比	月	オッズ比
1	2月	0.8844	2月	2.4407	12月	1.2094
2	1月	0.8534	1月	2.1587	1月	0.4514
3	12月	0.8133	12月	2.1336	2月	0.2806

表9 各月の重要POS大分類 (上位3位:ロジスティック回帰分析)

順位	12月		1月		2月	
	POS大分類	オッズ比	POS大分類	オッズ比	POS大分類	オッズ比
1	大分類10	1.6196	大分類12	2.2193	大分類16	23.6221
2	大分類9	0.9007	大分類13	1.7737	大分類10	1.2252
3	大分類7	0.8944	大分類10	1.2391	大分類6	1.0517

表10 各月の重要来店時間 (上位3位:ロジスティック回帰分析)

順位	12月		1月		2月	
	来店時間	オッズ比	来店時間	オッズ比	来店時間	オッズ比
1	時間20	1.6899	時間21	29.2813	時間12	1.2341
2	時間21	1.4763	時間13	1.2994	時間13	1.0344
3	時間14	1.3666	時間9	1.2666	時間9	0.9613

## 5. まとめ

本論文では小売店のPOSデータを対象とし、判定月より2カ月後の月の非継続来店顧客を検知するモデルを作成した。モデル化にはポテンシャル学習と呼ばれる新しいニューラルネットワークを使用し、平均予測正解率0.8344のモデル作成に成功した。また、ポテンシャル学習のもつ重要変数の抽出機能より、11時台から17時台の範囲で青果、加工食品、惣菜を購入する顧客が継続して来店している傾向にあり、加えて、各変数ジャンルで継続来店顧客と非継続来店顧客の占める比率はおおよそ80:20であることがわかった。消費者行動はポテンシャル学習を用いなければ明らかにならなかった事柄であり、非継続来店顧客の検知、そして消費者行動研究にポテンシャル学習が適していると判断できた。

本研究で使用したデータは、経営科学系研究部会連合協議会主催の平成26年度データ解析コンペティションで提供された全日食チェーン10店舗分のPOSデータのうち、ある1店舗についてのデータである。今後はほかの店舗でも同じ傾向があるのか、研究をさらに進めていく。



## 参考文献

- [1] R. Kamimura and R. Kitajima, "Selective potentiality maximization for input neuron selection in self-organizing maps," In *Proceedings of 2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2015.
- [2] R. Kamimura, "Self-organizing selective potentiality learning to detect important input neurons," In *Proceedings of 2015 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, to appear.
- [3] 市村直幸, 竹内俱佳, 永井豊, "パターン認識のためのクラスタ分析に基づく3層ニューラルネットワークの重み初期値の一設定法," 電子情報通信学会論文誌, **J77-D2**, pp. 301–310, 1994.
- [4] 下平丕作士, "ニューラルネットワークにおける誤差逆伝播法の学習性能向上のための重み初期値の設定方法," 情報処理学会論文誌, **35**, pp. 2046–2053, 1994.
- [5] 村田博士, 小野田崇, 由本勝久, 中野幸夫, 近藤修平, "建物の外から電気機器の使用実態を把握するモニタリングシステム—実家庭への適用実験—," 電気学会論文誌 C (電子・情報・システム部門誌), **124**, pp. 1874–1880, 2004.
- [6] 田辺和俊, 大森紀人, 小野修一郎, 鈴木孝弘, 松本高利, 長嶋雲兵, 上坂博亨, "ニューラルネットワークによる多種類の有機化合物の発ガン性の予測," *Journal of Computer Chemistry, Japan*, **4**, pp. 89–100, 2005.
- [7] R. Kamimura and R. Kitajima, "SOM knowledge induced learning with maximum information principle to improve multi-layered neural networks," In *Proceedings of the Sixth IASTED International Conference on Computational Intelligence (CI 2015)*, pp. 255–261, 2015.
- [8] 山下遙, 鈴木秀男, "セール品に注目した顧客の購買行動の解析—2値データのクラスタリングを考慮したロジスティック回帰分析—," オペレーションズ・リサーチ: 経営の科学, **60**, pp. 81–88, 2015.
- [9] Helsinki University of Technology (現: Aalto University), Laboratory of Computer and Information Science, "SOM Toolbox 2.0", <http://www.cis.hut.fi/somtoolbox/>