

東海大学大学院平成28年度博士論文

ニューラルネットワークを用いた
マーケティング分析に関する研究

指導 上村龍太郎教授

東海大学大学院総合理工学研究科

総合理工学専攻

北島 良三

目次

1	序論	9
1.1	マーケティングとは	10
1.2	マーケティング戦略	10
1.3	本論文で取り上げる「マーケティング」の対象	12
1.4	マーケティング分析	13
1.4.1	データマイニング	13
1.4.2	多変量解析	14
1.4.3	ニューラルネットワーク	15
1.5	マーケティング分析の現状	16
1.5.1	マーケティング分析とニューラルネットワーク	16
1.5.2	マーケティング分析と多変量解析	17
1.6	本論文の目的	19
1.6.1	ニューラルネットワークの脱ブラックボックス化	19
1.6.2	情報論的ニューラルネットワーク	20
1.6.3	マーケティング分析への特化	21
1.6.3.1	意見が異なるデータへの注目	21
1.6.3.2	注目レベルの調整	22
1.6.3.3	解釈可能	22
1.7	論文構成	23
2	解析手法	25
2.1	多変量解析	25
2.1.1	回帰分析	25
2.1.2	ロジスティック回帰分析	27
2.2	ニューラルネットワーク	28
2.2.1	ニューロンのモデル化	28
2.2.2	単純パーセプトロン	30
2.2.3	学習	31
2.2.4	多層パーセプトロン	32

2.2.5	誤差逆伝播法	33
2.2.5.1	数値を予測する場合	33
2.2.5.2	クラスを予測する場合	38
2.2.6	早期停止法	39
2.2.7	深層学習	39
2.2.8	自己組織化マップ	40
2.3	情報論的ニューラルネットワーク	43
2.3.1	情報理論	43
2.3.2	反復情報量最大化法	44
2.3.3	潜在学習法	47
2.3.4	反復潜在学習法	50
3	情報論的ニューラルネットワークの検証	53
3.1	解析データと解析手法	53
3.1.1	解析データ	53
3.1.2	解析手法	54
3.1.3	使用したニューラルネットワークの構造	56
3.2	検証結果	58
3.2.1	汎化能力検証	58
3.2.2	抽出重要変数の妥当性検証	59
3.2.3	学習第1段階が学習第2段階に与える効果の検証	65
3.2.4	変数着目効果の検証	66
3.2.5	初期重みをランダムに設定した場合の検証	68
3.3	ノイズ混入率が高いデータを用いた検証	72
3.4	多重共線性発生データに対する解析	79
3.5	構造学習との違い	85
4	消費者来店動機解析	91
4.1	解析概要	91
4.2	予測モデル概要	92
4.3	関東地域1店舗を対象とした解析	93

4.3.1	解析目的	93
4.3.2	解析データ	94
4.3.3	結果と考察	95
4.3.3.1	非継続来店顧客検知モデル	95
4.3.3.2	潜在学習法の潜在性の活用	99
4.3.3.3	潜在学習法を用いた消費者行動の解釈	101
4.3.3.4	従来手法を用いた消費者行動の解釈	103
4.3.4	まとめ	106
4.4	北海道地域5店舗を対象とした解析	108
4.4.1	解析目的	108
4.4.2	データ	108
4.4.3	モデル化手法	109
4.4.4	結果と考察	110
4.4.4.1	来店予測モデル	110
4.4.4.2	潜在学習法が獲得した知識について	112
4.4.4.3	潜在学習法による消費者行動解釈	114
4.4.4.4	ロジスティック回帰分析による消費者行動 解釈	117
4.4.5	まとめ	119
4.5	全店舗を対象とした解析	121
4.5.1	解析目的	121
4.5.2	解析データ	121
4.5.3	解析手法	121
4.5.4	結果と考察	122
4.5.4.1	深層学習	122
4.5.4.2	事前学習併用誤差逆伝播法	123
4.5.4.3	誤差逆伝播法	124
4.5.4.4	重要変数解釈	124
4.5.5	まとめ	127

5	履修科目決定行動解析	131
5.1	解析目的	131
5.2	精緻化見込みモデル	132
5.3	消費者行動モデルと学生履修行動	132
5.4	調査方法と解析方法	134
5.4.1	アンケート調査	134
5.4.2	解析方法	134
5.5	結果と考察	136
5.5.1	アンケート調査結果	136
5.5.2	データ解析結果	136
5.5.3	変数解釈結果	137
5.6	まとめ	137
6	重要 Tweet 判断基準抽出	139
6.1	解析概要	139
6.2	災害と Twitter	139
6.3	モデル作成概要	142
6.4	Tweet 分類モデル作成検証	143
6.4.1	解析目的	143
6.4.2	Tweet データの形態素解析	143
6.4.3	解析用データの作成	143
6.4.4	Tweet 分類モデル	144
6.4.5	分類結果	147
6.4.6	まとめ	149
6.5	重要判断基準抽出	150
6.5.1	解析目的	150
6.5.2	解析データ	150
6.5.3	結果評価尺度	151
6.5.4	結果と考察	152
6.5.5	まとめ	154

7	企業行動原理抽出	157
7.1	解析概要	157
7.2	社是のデータ化	158
7.3	東証二部上場の4業種を対象とする解析	160
7.3.1	解析目的	160
7.3.2	分析	161
7.3.3	結果と考察	162
7.3.3.1	業種：機械	162
7.3.3.2	業種：金属製品	163
7.3.3.3	業種：建設業	163
7.3.3.4	業種：不動産業	164
7.3.3.5	各グループと収益	164
7.3.4	まとめ	165
7.4	東証一部製造業企業を対象とする解析	171
7.4.1	解析目的	171
7.4.2	データ	171
7.4.2.1	解析対象社是の収集	171
7.4.2.2	解析用データの前加工	172
7.4.2.3	収益性データ	173
7.4.3	解析手法	173
7.4.4	結果と考察	174
7.4.4.1	SPLによる社是の分類の結果	174
7.4.4.2	SPLによる分類結果の要因	175
7.4.4.3	SPLによるターゲットの予測結果	175
7.4.4.4	分類結果の構造	176
7.4.4.5	分類結果と重要変数との関係	177
7.4.4.6	分類結果と収益性との関係	177
7.4.5	まとめ	178
7.5	東証二部製造業企業を対象とする解析	181
7.5.1	解析目的	181

7.5.2	解析対象企業と解析データ	181
7.5.3	解析手法	183
7.5.4	解析手順	183
7.5.5	結果と考察	184
	7.5.5.1 社是分類モデル	184
	7.5.5.2 モデル解釈	185
7.5.6	まとめ	187
8	各解析手法まとめ	189
8.1	各解析結果	189
	8.1.1 ロジスティック回帰分析による解析結果まとめ . . .	190
	8.1.2 誤差逆伝播法による解析結果まとめ	193
	8.1.3 深層学習による解析結果まとめ	194
	8.1.4 提案手法による解析まとめ	195
	8.1.4.1 潜在学習法	195
	8.1.4.2 反復潜在学習法	197
8.2	提案手法の限界点	198
9	結論	207
9.1	SWOT 分析	207
9.2	潜在学習法のポジショニング	211
	謝辞	215
	参考文献	217
	業績一覧	229

1 序論

本論文はニューラルネットワークがマーケティング分析の解析手法として一般的に選択されていない主たる原因が、ニューラルネットワークの持つブラックボックス問題 (低モデル解釈性) であるにとらえ、ニューラルネットワークの脱ブラックボックス化について検証し、脱ブラックボックス化ニューラルネットワークのマーケティング分析への適用可能性について論じたものである。一般的にマーケティング分析では解析手法として多変量解析が選択されている。しかし、多変量解析は複雑なデータ解析を得意とはしておらず、また、多重共線性が発生したデータに解析作業・結果が左右されるという問題が存在している。一方、ニューラルネットワークは一般的な統計手法と比較して複雑なデータ解析が可能であるという特徴を有しており、また、多重共線性問題から影響を受けにくい。そのため、ニューラルネットワークは消費・製品・流通などという複雑な事象を解析対象とするマーケティング分析において有効な解析手法であると考えられる。ところがニューラルネットワークにはブラックボックスと呼ばれるモデル解釈 (内部表現解釈) が困難であるという問題が存在している。モデル解釈はマーケティング分析では重要な事柄であるため、このブラックボックス問題の存在はマーケティング分析において問題である。しかし、ニューラルネットワークは前述したように複雑なデータの解析が可能であり、多重共線性問題もあまり問題にならないため、脱ブラックボックス化に成功した (高モデル解釈性を有する) ニューラルネットワークが提案できればそれは新たなマーケティング上の発見へとつながる強力な解析手法になると考えられる。本論文ではこのニューラルネットワークの脱ブラックボックス化とマーケティング分析への応用について論じていく。本論文の序論である本章では、研究対象となるマーケティング、ニューラルネットワーク、そして、これらを取り巻く現状について述べ、本研究の目標について整理を行う。

1.1 マーケティングとは

日本マーケティング協会の定義 [1] によるとマーケティングとは、「企業及び他の組織がグローバルな視野に立ち、顧客との相互理解を得ながら、公正な競争を通じて行う市場創造のための総合的活動」である。市場を創造するという最終目標はあるものの、活動内容は「総合的」とあるように幅が広いことを示している。これはすなわち「市場創造へ繋がるのであれば、どのような活動もマーケティングである」と言えるものである。

このようにさまざまな活動が考えられるマーケティングであるが、製品についての活動 (product)、価格についての活動 (price)、広告や販売促進についての活動 (promotion)、流通についての活動 (place) が一般的であろう。これらは Edmund Jerome McCarthy によって提唱された 4P と呼ばれる分類である [2]。マッカーシーによれば 4P は数多く存在するマーケティング要素を切り詰めた結果であるという。今日この 4P はマーケティングを考える上での基本事項となっている [3] [4]。なおこれは広告の視点からであるが、例えば広告や販売促進はさらに、販売促進 (sales promotion)、人的販売 (personal selling)、商品パブリシティ (product publicity)、商品広告 (product advertising)、という「マーケティング・コミュニケーション」 [5] と呼ばれるコミュニケーションを軸としたマーケティングの世界へと繋がっており、このことはマーケティング要素の膨大さの一例を示すものである。

1.2 マーケティング戦略

マーケティングで最も大切な事柄は、マーケティング戦略の立案である。どのような市場で、どのような製品を、誰に、いくらで、どのように供給し、どのように販売するのか、という全体的なコンセプトを綿密に設計する必要があるのである。マーケティングを成功させるためには各領域でのマーケティング活動を密接に関連させ、全体的なもの (戦略) として考える必要がある。

マーケティングは幅広い活動領域を持つため、全活動領域でのマーケティング活動を設計することは容易ではない。しかし、だからといって各領域で単発的にマーケティングを実施してもその効果を最大限に発揮することはできない。製品は製造されてから流通網に乗って販売される場所へと届けられ、最終的に消費者に届き、そして消費される。この流れより、例えば価格だけに着目してマーケティングを実施しても全体としてのマーケティングを満足させることができないことは明らかである。非常にリーズナブルな価格が設定された製品も流通が機能せずに売り場に届かなければ販売していないのと同じであるし、また、製品の内容が販売対象とする消費者像と乖離していた場合、価格が手頃であっても購入されることはないだろう。

マーケティングの全体 (戦略) を設計するのに効果を発揮するのが、先にあげたマッカーシーの 4P である。マーケティング戦略は和田らによると「市場需要の創造・開拓・拡大を目的としてターゲットを設定し、それに対応したマーケティング・ミックス要素を設計することによって構造化される」[6] ものである。ターゲットおよびマーケティング・ミックスとは、前者が製品販売対象のこと、後者がさまざまなマーケティング活動のことである。このマーケティング・ミックスを集約するとマッカーシーの 4P (「Product」、「Price」、「Promotion」、「Place」) となるのである。

マーケティング戦略を設計する上で最重要事項は目標の設定である。マーケティングにより実施する事柄が明確になっていないと、効果あるマーケティングは実施できない。この目標設定で定める対象市場は、新規市場の場合は「創造」であり、既存市場が対象になる場合は「拡大」である。いずれにせよ、どの市場を目標とするのかを明確にすることが大切である。続いて、その市場の誰を対象とするのかという事柄を決定する。市場にはさまざまな性別・年齢・職業・価値観などを持った顧客が存在しているので、市場創造あるいは市場拡大を目標として誰に働きかけるのかを決定するのである。市場と顧客が決定すると戦略の設計はより詳細な事柄に移る。ここで登場するのが 4P である。目標市場と顧客に訴えかけるための商品設計 (product)、いくらで販売するのかの価格設計 (price)、

どこで販売するのかの流通設計 (place)、製品の存在をどのように告知するのか、どのような販売方法をとるのかのプロモーション (promotion)、と各分野を詳細に設計し、最終ゴールへ向けてそれぞれを調整する。これがマーケティング戦略である。消費者からすると新商品が突然店舗に登場する印象を受けるが、その裏には多岐に渡る綿密な設計がされたマーケティングが隠れているのである。

1.3 本論文で取り上げる「マーケティング」の対象

ここで本稿で述べるマーケティングの対象について述べる。本論文ではマーケティングの領域のうち主に product の領域に焦点をあてる。特に「製品がなぜ売れるのか?」、「どのような消費者に購買されているのか?」、「消費者はどのような購買行動を行うのだろうか?」という消費者に関連する点を対象とする。

このように本論文ではマーケティングのうち消費者行動に関する事柄を研究対象としている。しかし本来、専門的に消費者行動を研究対象とする学問は「消費者行動研究」である。ここで、マーケティング研究として消費者行動研究を取り扱うことについて触れておきたい。消費者行動研究の目的は井上 [7] によると「消費者はどのようなプロセスを経過して購買に至るのか」、「消費者の購買・消費の経験がどのように蓄積され、次の購買に影響していくか」、「そのプロセスの中で、どのような対象 (人・組織) からどのような影響を受けているのか」、「その購買の結果あるいは使用にどの程度満足しているのか」、「マーケティング諸活動がその購買決定に対してどのような影響を与えることができるのか」、「消費者の行動がマーケティング諸活動に対してどのような影響をあたえるのか」に整理できるという。

消費者行動研究の研究目的に「マーケティング」という単語が出現していることから明らかなように、また、マーケティングが市場で顧客 (消費者) に商品を販売する仕組み設計であることから見ても、消費者行動研究とマーケティングは互いに密接に関連していることが分かる。

1.4 マーケティング分析

1.4.1 データマイニング

本論文ではデータ解析を通して「消費者はどのような購買行動を行うのだろうか?」といったマーケティング上の課題についてアプローチを行う。マーケティング施策を策定するにあたり施策の根拠は重要である。根拠なき施策は説得力を持たず、また、実施すべきかどうか客観的に判断を下せないからである。実施の根拠となりえる要素は、消費者の声、小売店の反応、自社の置かれた環境などさまざまあり、施策立案にはこれらについて情報を収集する必要がある。

収集した情報はそのままでは使用できないため、「まとめる」、あるいは「加工」することが必要である。情報を収集しても、要約や集計などの整理をしておかないと、それは情報が使用できる状態にあるとは言えないし、また、収集した情報をさらなる分析にかけることもできない。

収集した情報は集計からの現状把握の他「なぜそのような回答を消費者はしたのか」といった回答の背後を調査することにも使用できる。例えば賃貸住宅についてある地域を調査したと仮定する。調査項目を最寄駅からの距離、築年数、広さ、近隣商施設の種類、家賃とした場合、集計結果からは各調査項目の平均値、標準偏差、最大値、最小値などが得られるが、分析をすすめるとそれぞれの調査項目の関係が明らかになり、さらにすすめると、家賃に影響を与えている項目の抽出が可能となる。また、この段階まで分析がすすむと、各項目を変化させ、その際の家賃の予測を行うシミュレーションが実施できる。これが可能なのは各項目から関係(特徴)が抽出されたからであり、このようにデータから特徴を抽出する分析手法を「データマイニング」という。

データマイニングは英語の「Data」と「Mining」を合わせた造語であり「データを発掘する」という意味を表す。つまり「データ」という山を発掘し、「特徴」という鉱物を掘り出すことを目的としている。特徴とは出力に影響を与える入力のことであり、マーケティング効果と施策の関係を説明するものである。

マーケティングは実際に実施してみなければ効果が未知数である。なぜならマーケティングとは 1.1 で述べたように「さまざまなマーケティング・ミックスが行われ、その結果、消費者が刺激され行動を行う」からであり、消費者に作用する要素数が多く、また消費者の反応も各自まちまちであるからである。マーケティング・ミックスはマーケターがコントロールするものの、消費者でどのように処理されるかは誰にも分からない。よってマーケターが分かるのは実施した内容とその結果だけである。すなわち作用の過程が見えないのである。しかし過程が見えないからといって、過程を検証しないわけにはいかない。それはマーケティングの効果を把握するためであり、分析依頼者(クライアント)に説明するためである [8]。よってマーケターは施策実施内容という「入力」と、実施結果という「出力」から作用過程を解釈しなければならない。この時活用できる方法がデータマイニングである。

1.4.2 多変量解析

データマイニングを実施するための手法は多数存在しているが、マーケティング分析で一般的に使用されるのは統計学に基づいた「多変量解析」である [8]。多変量解析とは、例えばマーケティング効果を y 、 k 番目のマーケティング施策を x_k ($k = 1, 2, \dots, K$)、パラメータを a_k ($k = 1, 2, \dots, K$) と b で表し、この y と x_k の関係を $y = a_1x_1 + a_2x_2 + \dots + b$ で表現して解析する手法のことである。式中のパラメータ値を調べることでマーケティング施策の効果に対する影響を把握することができるのである。

しかし分析(予測・分類)の精度といった観点から見ると後述するニューラルネットワークといった手法には及ばない点もあり、一般的に使用されているとは言え、多変量解析を用いることがマーケティング分析の最良の選択であるということとはできない。

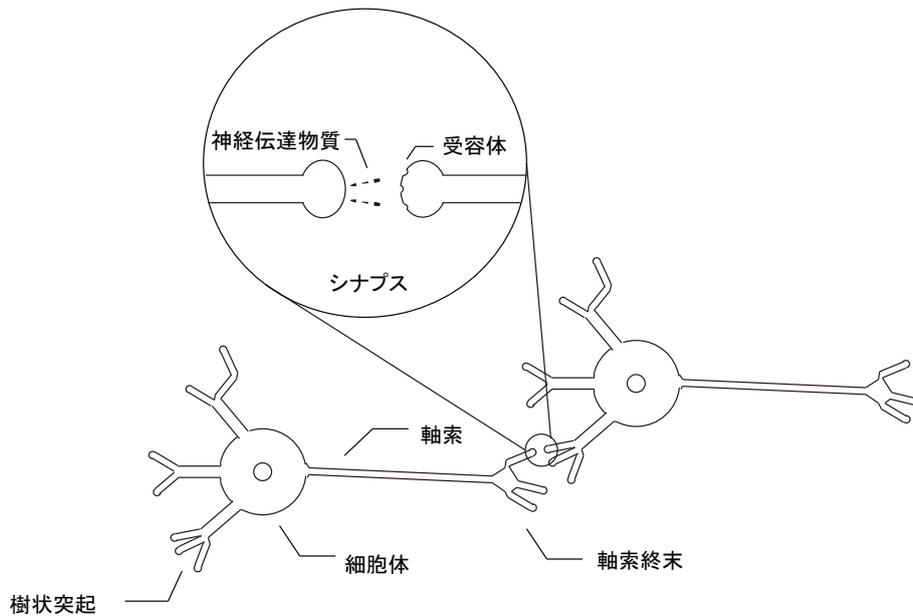


図 1: 神経細胞の構造図

1.4.3 ニューラルネットワーク

データマイニング手法には統計学とは異なるアプローチで実施する手法もある。例えば人間の脳の情報処理を基とする「ニューラルネットワーク」である。

ニューラルネットワークとは「ニューロン」 [9] のネットワークのことを意味しており、ニューロンとは神経細胞を意味する。ニューロンは図 1 に示す構造をしている。簡単に構造について述べるとニューロンは、その本体である「細胞体」、他のニューロンから刺激を受け取る役割を持つ「樹状突起」、他のニューロンへ刺激を伝達する役割を持つ「軸索」と「軸索終末」から構成されている。ニューロンは刺激を受け取ると情報処理を行い結果を出力するという動作を行う。このニューロンが互いに結びついてネットワークを構成し、刺激が次から次へと受け渡されて神経系を巡っていく。また、ニューロン間の接続はシナプスという構造を介して行われる。このシナプスでは情報送信側のニューロンの軸索終末から神経伝達物質と呼ばれる化学物質が放出され、情報受信側で樹状突起にある受容体がこの神経伝達物質を受け取りニューロン間での情報伝達が行われる。これが生体でのニューラルネットワークの大まかな流れであり、このしくみを模したものが分析手法のニューラルネットワークである。

1.5 マーケティング分析の現状

1.5.1 マーケティング分析とニューラルネットワーク

ニューラルネットワークは多変量解析と比較して、高精度なデータ予測・分類が可能という特徴を持っている。ニューラルネットワークはその高精度なデータ予測・分類機能を活かし、理工学分野、化学分野、電磁気学分野、金融分野などさまざまな分野で活用されている [10], [11], [12], [13]。もちろん、ニューラルネットワークはマーケティング分野でも使用されている [14]。

しかし、1.4.2 でも述べたが中心的な手法とはなっていないのが現状である [15], [16], [17]。これは主にニューラルネットワークの持つ「ブラックボックス」という問題点からである [18]。ブラックボックスはニューラルネットワークから出力される結果の解釈が困難であるさまを表現している [19]。結果の解釈とはすなわち入力と出力の関係を明らかにすることであり、1.4.1 でマーケティングは過程が見えないと述べたが、この過程を調べる (解釈する) ための手法自体にもブラックボックスと呼ばれる「過程が見えない問題」が存在しているのである。

マーケティングの効果は説明できる必要があることを 1.4.1 で述べたが、このブラックボックスの存在がニューラルネットワークがマーケティング解析で中心的に使用されない理由と考えられる [20], [21]。1.4.2 で述べたように多変量解析では、パラメータ値を解釈することによりマーケティング施策の消費者での処理過程解釈が容易にできるため、解釈可否の差は大きいものである。例えシミュレーション精度が多変量解析よりもニューラルネットワークの方が高いとしても、「何故その結果が得られたのか」が解釈できない限りは解析手法として採用することに躊躇してしまう。また信頼性という観点から見ても、解釈可能な手法の方が信頼 (納得) できるため、分析手法として採用を決めやすいだろう。

解釈は第三者に説明する際に特に重要性を持つ。解析した結果得られる特徴群 (モデル) を解析者自身が使用する分には解析者本人が納得しさえすれば良いが、第三者より解析依頼を受けてモデル作成を行う場合、「な

ぜか不明だが良い結果が出力される」という解析報告と共に納品されるモデルでは説得力に乏しく、また第三者からのそのモデルに対する信頼性は低くなる。実務の分野では解釈可能な手法が好まれて使用されるが [22]、このことは説得力が影響していると考えられる。

1.5.2 マーケティング分析と多変量解析

1.5.1 で触れたようにニューラルネットワークはブラックボックスを抱えており、多変量解析と比較すると説明力が高くないという特性を持っている。ではこのことをもって「マーケティング分析には多変量解析が最適なデータマイニング手法である」といえるのだろうか。

データマイニングを実施するためにはデータを収集し、解析を行わなければならない。データ収集手段は、消費者に対するインタビューやアンケートの実施といった直接的な収集方法、POS データといった記録から購買行動を抽出する間接的な収集方法があげられる。直接的・間接的と表現したが、必ずしも直接的な収集方法により得られたデータが優れているというものではない。人間の行動は複雑であり行動者自身が行動理由を把握していない場合もあるからである。また、インタビューの質問の方法などによっても得られる結果が左右され、直接的手法はデータ収集が容易な方法とは言えない。一方、間接的手法は行動結果のデータであり、また、収集は容易である。しかし、その結果に至った過程はデータに含まれないため、間接的手法により収集されたデータから例えば購買行動を得ることは、点と点を線で結ぶような作業といえる。しかも線は明確に引けるものではなく分析手法による推測によらねばならない。そのため解析には十分な数のデータサンプルと、解析手法にはデータからより多くの特徴を抽出できる手法が必要となる。

多変量解析手法は統計的なバックボーンを持つデータ解析手法であり、マーケティング分析で一般的に利用されている解析手法である。しかし解析できるデータに制限も多く、解析者の思い通りにデータを解析するという点では課題が存在していると言える。

解析上の制限とは例えば、近い関係にあるデータ (変数) を用いて解析すると結果に悪影響を与えるといった事柄である [23]。これは「多重共線性」と称される問題である。解析者から見るとデータの傾向が似てしまうことは、あえて意識して同じような傾向のデータを収集しない限りは結果的に生じることであり、多重共線性の発生回避は容易なことではない。また、似通っていたとしてもデータ収集元が異なるなど、収集した意図が異なるものでもある。

多変量解析では多重共線性の発生しているデータを処理すると、「結果が求められない」、「結果の解釈ができない」などといった不都合が生じてしまう [24]。そこで、このような事態を避けるためデータ (変数) を削減する必要がある。仮説を立ててデータ収集を行い集めた解析データを、解析する段階で分析上の制限という理由により変数削減する必要があることは「データを用いて解析対象をモデル化しようとしているのか」、「モデル化手法に解析対象を合わせているのか」という疑問を生じさせる。もちろん、計算コストの削減や、解釈のし易いモデルの作成という観点からの変数削減は有益であるので変数削減そのものの是非は否定しない。ここで述べていることは手法上の制限により止むを得ず行われる変数削減についてである。

このような解析上の制限が存在しているものの、マーケティング分析では多変量解析が解析手法として採用されることが多い。何度も述べているがこれは、データ予測や分類の過程が説明できるということがその理由である。マーケティング分析の結果はマーケティング上の意思決定に利用される。マーケティングは活動に多くの資金が必要 (1970 年代時点でのことではあるが消費者が支払う代金の 50%前後がマーケティング費用だと言われている [2]) となるばかりではなく消費者の意識にまで影響力を与えるものである。そのため間違った意思決定を行った結果、消費者が離れてしまうなどという事態も起こりえる。この場合、顧客を他社から取り戻すことは容易ではない。このことから、データ解析の結果が十分に説明でき、かつ、納得できるものでなければ、解析結果を意思決定に採用することは大きなリスクとなるのである。

なお、ニューラルネットワークは多重共線性が発生しているデータの解析が困難であるというような制約を特に持たないため [25]、変数を削減するという作業は必要なく、解析者の自由度が高い手法と言える。このことは多変量解析と比べて利点である。

1.6 本論文の目的

1.6.1 ニューラルネットワークの脱ブラックボックス化

これまで述べてきたように、ニューラルネットワークをマーケティング分析に適応する場合、たしかに説明力という点では多変量解析と比べて差がついてしまう。しかし予測・分類精度、解析上の制限という点を考えるとニューラルネットワークは大きな可能性を持っている。鍵はニューラルネットワークの「脱ブラックボックス」である。

脱ブラックボックスが可能となればニューラルネットワークは、高精度な結果を出力する、解析者にとって使用勝手がよい、そして依頼者にとって信用が高いデータ分析手法と成りえると考えられる。このことについて検証を行うのが本論文の目的である。

図 2 は本論文で目指す脱ブラックボックス化ニューラルネットワーク手法の位置づけについて示したポジショニングマップである。ポジショニングマップは定めた軸を基に相対的な比較を行うフレームワークであり、ここでは軸を「解析データの複雑さ」と「解釈性」の 2 軸とした。解析データの複雑さとは本論文では「多重共線性が疑われるデータ」や「線形モデルで対応できないデータ」を意味する。多変量解析は複雑なデータ解析に向いていない手法であるが、解釈性は高い手法であるので複雑さが小さく、解釈性が高い右下に位置する。そしてニューラルネットワークは複雑なデータ解析が可能なものの、解釈性は低い手法であるので左上に位置する。この二つの既存手法に対して提案手法は解釈が容易であり複雑なデータ解析ができるもの、すなわち、ニューラルネットワークを「解釈が容易」である方向へシフトさせることを試みる。

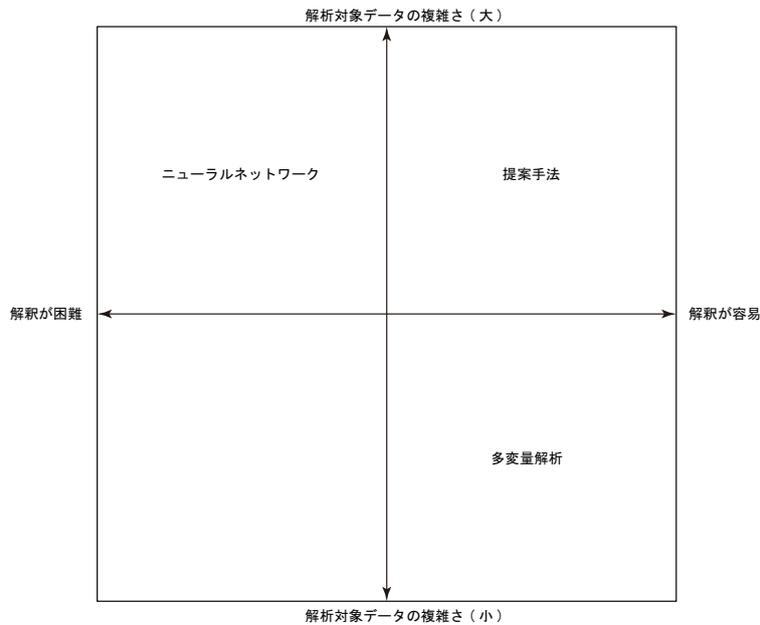


図 2: 多変量解析と提案手法のポジション

1.6.2 情報論的ニューラルネットワーク

ニューラルネットワークがブラックボックス化してしまう理由として多くのニューロンが活動することがあげられる。ニューラルネットワークは多くのニューロンと層により複雑に情報処理される手法であるので、多くのニューロンが情報処理に関係することはニューラルネットワークの特徴であるが、このことが反面、入力と出力の関係性解釈を困難にさせている。これを例えば少数のニューロンのみを情報処理に使用するという方法により解消し、入力と出力の関係解釈を容易にすることが、本論文で目指すニューラルネットワークの脱ブラックボックスである。ネットワークで活動するニューロンを絞る手法はこれまでも提案・検証されている [26], [27], [28], [29], [30], [31], [32]。しかし、膨大な計算時間が必要、学習と絞り込みの交互実施が必要、といった複雑さがあり、よって本論文では従来手法よりもより平易なメカニズムにて脱ブラックボックスを目指す。詳細は次章で述べるが、本論文では分散を基 [33] とする情報論的ニューラルネットワークを使用する。

解析手法が平易であることはモデルの解釈性に次いで大切なことである。なぜならデータ解析の専門家ではない第三者に解析結果を説明する

場合、解析手法が平易で理解が容易な方がより説得力が高まるからである。専門家ではない第三者に対して詳細な専門的説明は通常要求されないが、動作原理の概要を説明することは必要で、この際に手法の平易さが生きてくる。

またこの「分散に着目する」ことであるが手法を平易なものとする以外に、マーケティング分析に特化したニューラルネットワーク手法とするために用いる。本論文で提案する手法には以下三つの機能を有しておりマーケティング分析向けの解析手法となっている。

1.6.3 マーケティング分析への特化

1.6.3.1 意見が異なるデータへの注目

マーケティング分析で取り扱うデータを整理すると、「各人で意見が分かれるデータ」と「各人で意見が分かれなデータ」の二つに分類が可能である。前者は配色・デザイン・機能などに対する消費者の意見、後者は生理的あるいは安全などに対する欲求などが該当する。このうち「意見が分かれなデータ」は調査対象が多数存在していても回答は同様であり(例えば「安全に暮らしたいですか?」と問いかけた場合「はい」と回答する人が多数であろう)、このデータは解析をするまでもなく傾向が把握できるものである。よって傾向が容易に捉えられないデータこそがマーケティング分析で主として解析したい事柄であり、重要なデータである。本論文では分散に注目して学習を行う情報論的ニューラルネットワークを提案するが、意見が異なるデータとは分散が大きいデータであるため、この動作は「意見が異なるデータに注目して学習を行う」と言い換えることができる。これはすなわち、マーケティング分析で重要なデータに注目をするという意味である。

この「注目する」という動作は消費者の購買時の意思決定とも関連がある。我々は製品購入時に「予算」、「デザイン」あるいは「機能」といった事柄に注目し、購入候補を絞り込んで最終決定判断を下すが、この「何かに注目する」という購買時の行動が提案手法に組み込まれていると言

える。

1.6.3.2 注目レベルの調整

上記の「意見が異なるデータへの注目」であるが、提案手法ではこの注目レベルを調整することが可能である。先に上げた「予算」、「デザイン」あるいは「機能」の三つの例を再び利用するが、注目レベルの調整とは、例えば購買時に「デザインを最も重視して次に予算を重要視して」などと購買候補を絞り込むようなことである。すなわち、注目するデータを注目度合に応じて差をつけるのである。

注目レベルを上げることにより、ある変数が極端に強調されてしまうことになるが、マーケティング分析においてこの強調は有効である。目的変数への貢献度合いが不明な多数の変数から特徴を見つけ出すのがデータマイニング手法を用いたマーケティング分析であるが、これら変数のなかから注目する変数を選択し（糸口を見つける）、さらにそれを強調することで、全変数を均等に取り扱っては見えなかった事柄の抽出が期待できるからである。このため、従来手法では予測・分類精度が上がらなかったデータの解析や抽出できなかつた特徴を得ることが提案手法による解析では期待できる。

1.6.3.3 解釈可能

マーケティング分析では解釈が重要であるとこれまで述べてきた。そのため提案手法はモデル解釈が可能という機能を有している。提案手法は「意見が異なるデータへ注目」し「注目レベルを調整」して学習を行うが、「注目した変数」と「どのくらい注目したのか」を調べることで、解析手法が何をどのくらい注目して解析したのか、すなわち、モデルの解釈が可能となる。

本論文ではこのマーケティング分析に特化した新しい解析手法を予測・分類精度、解釈性、解析特性などから検証し、マーケティング分析への応用について論じていく。

1.7 論文構成

本論文は合計9章から構成されている。1章ではマーケティングとマーケティング分析をとりまく問題、そして本論文の目的について分析手法の観点から述べた。2章以降では1.6.1で述べた「脱ブラックボックス化ニューラルネットワーク手法」と多変量解析との分析結果の違いについて述べていく。

具体的には2章では分析手法について、多変量解析・ニューラルネットワーク・情報論的ニューラルネットワークについて述べ、3章では本論文でマーケティング分析への適用を提案する情報論的ニューラルネットワークのうち、主となる「潜在学習法」についてモデル汎化能力・抽出重要変数の妥当性・多重共線性が発生している場合の解析などを検証する。そして4章では、POSデータを用いた消費者行動についての解析、5章では大学生が履修授業を選択する際の行動についての解析、6章では災害発生時のTweet情報を対象とした「重要情報・非重要情報の判定」についての解析、7章では、社是を対象とした企業行動についての解析について述べる。そして8章にて3章、4章、5章、6章、7章の総括を行い、9章で結論を述べる流れである。

2 解析手法

本章では4章、5章、6章、7章で述べる各解析で使用する手法について、大きく分けて「多変量解析」、「ニューラルネットワーク」、「情報論的ニューラルネットワーク」の3手法について述べる。具体的には多変量解析では「回帰分析」と「ロジスティック回帰分析」について述べ、ニューラルネットワークでは「単純パーセプトロン」、「多層パーセプトロン」「自己組織化マップ」、情報論的ニューラルネットワークでは「反復情報量最大化法」、「潜在学習法」、「反復潜在学習法」について合計8手法について述べる。

2.1 多変量解析

2.1.1 回帰分析

マーケティング分析を実施する分析手法のなかで主流となっている手法は多変量解析などの統計的手法である [8]。数多くの手法が存在しているが、本章では代表的手法である「回帰分析」と「ロジスティック回帰分析」を取り上げることとし、はじめに回帰分析から述べる。

回帰分析 [34], [35] はデータを $y = a + b_1x_1 + b_2x_2 + \dots + b_kx_k (k = 1, 2, 3, \dots, n)$ で近似して表す分析方法である。 x は変数、 y は変数が作用した結果、また、 a と b_k はパラメータを表す。例えば、気温を変数として清涼飲料の売上個数を分析する場合、気温が x 、売上個数が y である。ここで、 x を説明変数、 y を目的変数と称する。図 3 (a) は気温と清涼飲料の売上個数を示したデータ (架空データ) を散布図にしたものであり、この散布図に、最小二乗法と呼ばれる手法 (後述する) を用いて直線 (回帰直線) を引いたものが図 3 (b) である。この直線は図 3 (b) の場合 $y = 1.257x - 24.344$ で示され、この式を用いることにより y の値を x の値から算出することができる。このようにデータを近似する直線式 (回帰式) を求め分析を行う手法を回帰分析といい、データの近似にくわえ予測も可能な解析手法である。

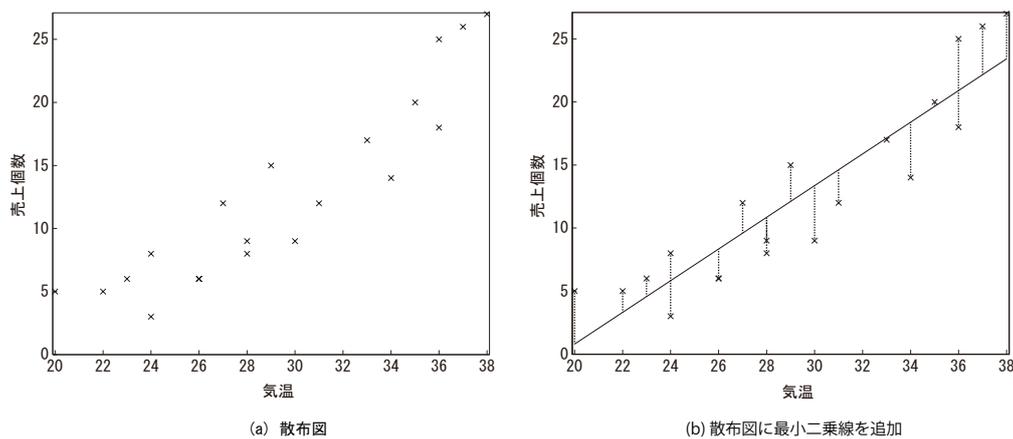


図 3: 散布図

最小二乗法は i 番目の y の値 y_i と近似されて得た値との差 (誤差) が最も小さくなるようにする方法である。誤差はデータのサンプル番号を i 、 k を変数番号、 y_i と $x_{k,i}$ をそれぞれデータサンプル i 番目の目的変数と変数番号 k の i 番目サンプルの説明変数とすると

$$\varepsilon_i = y_i - (a + b_1x_{1,i} + b_2x_{2,i} + \cdots + b_kx_{k,i}) \quad (2-1)$$

で求められ、これの平方和は

$$S = \sum_{i=1}^I \varepsilon_i^2 \quad (2-2)$$

により計算される。ここで I はサンプル個数である。すなわち最小二乗法とは誤差の平方和が最小になるパラメータ a と b を求める手法である。

パラメータ a と b_k は、 S を a および b_k でそれぞれ偏微分して 0 とおいた連立方程式 $\partial S/\partial a = 0, \partial S/\partial b_1 = 0, \cdots, \partial S/\partial b_k = 0$ を解くことにより求められる。

回帰分析はシンプルなモデルではあるが、データの近似・予測に加えてパラメータについて確認することで、説明変数が目的変数に与えている影響を解釈することができる手法である。

2.1.2 ロジスティック回帰分析

2.1.1 で述べた回帰分析の場合、目的変数には実数を使用されるが、分析によっては目的変数が「0」か「1」、すなわち「起こらない」、「起こる」のような2値しか取らない場合もある。2値しかとらない値に対して回帰分析は使用できないため、このような目的変数を用いる場合、ロジスティック回帰分析 [24] が用いられる。

ロジスティック回帰分析が扱う目的変数は「起こらない・起こる」であり、これはすなわち確率である。ここで「1を起こる」、「0を起こらない」とし、 i 番目のサンプルの目的変数 y_i の確率を $P(y_i = 1) = \theta$ 、 $P(y_i = 0) = 1 - \theta$ とおく。 θ は成功率を、 $1 - \theta$ は失敗率を意味する。この成功率の対数オッズ ($\log \{\theta / (1 - \theta)\}$) は $-\infty$ から ∞ の値をとるため、回帰分析で扱えるようになる。この変換をロジット (logit) といい

$$\text{logit}(\theta) = \log \frac{\theta}{1 - \theta} \quad (2-3)$$

で表される。

ロジスティック回帰分析は a と b をパラメータ、 x を説明変数、 $k(= 1, 2, \dots, K)$ を変数番号、 $i(= 1, 2, \dots, n)$ をデータサンプル番号とすると

$$\text{logit}(\theta_i) = a + b_1 x_{1,i} + \dots + b_k x_{k,i} \quad (2-4)$$

にあてはめるものである。これは θ について解くと

$$\theta_i = \frac{1}{1 + e^{-(a + b_1 x_{1,i} + \dots + b_k x_{k,i})}} \quad (2-5)$$

である。この式について a と b を求めることによりロジスティック回帰分析が実施される。これは解析的に得ることは困難であり、数値計算により求められる。

ロジスティック回帰分析も回帰分析と同様にデータの近似・予測に加えてパラメータについて確認することにより、説明変数が目的変数に与えている影響を解釈することができる。

2.2 ニューラルネットワーク

2.2.1 ニューロンのモデル化

ニューラルネットワークは 1.4.3 で述べたように人間の脳で行われる情報処理を模した手法である (図 1)。1.4.3 では生体のニューロンについて述べたが、ここではデータ解析手法のニューラルネットワークを構築するために、ニューロンを模した人工的なニューロンについて述べる。

図 4 に人工ニューロンを示す。これは「形式ニューロン」と呼ばれる Warren S. McCulloch と Walter Pitts によって 1943 年に発表されたものである [9]。図中の x_i ($i = 1, 2, \dots, k$) は他のニューロンからの入力を意味しており、ニューロンとの接続部分はシナプスを模したものである。ここに w_i ($i = 1, 2, \dots, k$) とあるが、これは情報の伝わりやすさを意味する「重み」と呼ばれている値であり、他のニューロンから伝達されてきた情報の伝達量をコントロールする機能を持つ。

重みを介してニューロンに伝わった情報は総和をとられ、ニューロンへの入力 u となり、関数 $f(u)$ で処理される。この総和と関数での処理を 1.4.3 では「刺激を受け取ると情報処理を行う」と表現していた (人工ニューロンと生体のニューロンとの対比を図 5 に示す)。

ここで使用する関数はさまざまな種類があるが、ヘヴィサイド関数 (図

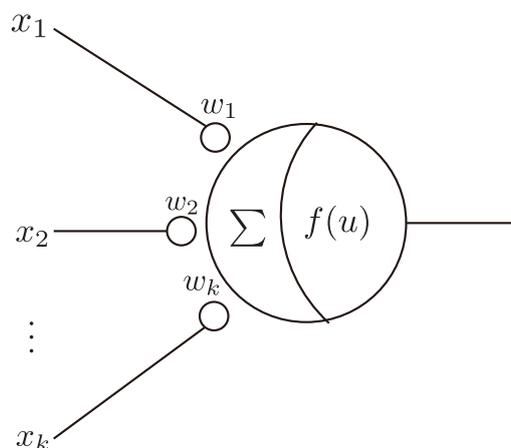


図 4: 形式ニューロンモデル

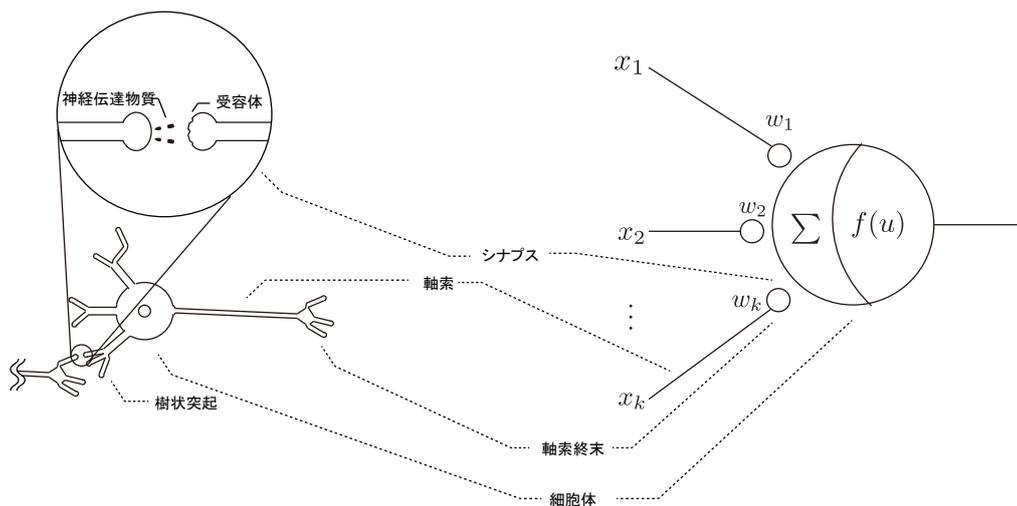


図 5: ニューロンと形式ニューロンの比較

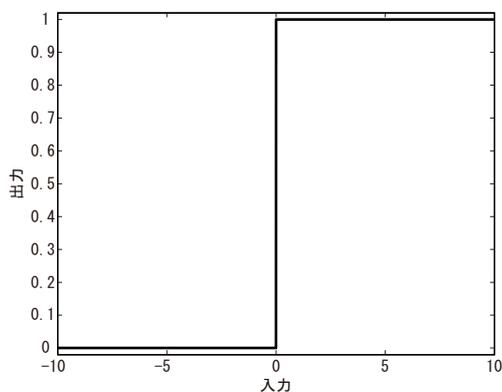


図 6: ヘヴィサイド関数

6)

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2-6)$$

や、標準シグモイド関数 (図 7)

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2-7)$$

および双曲線正接関数 (図 8)

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2-8)$$

が一般的である [36]。

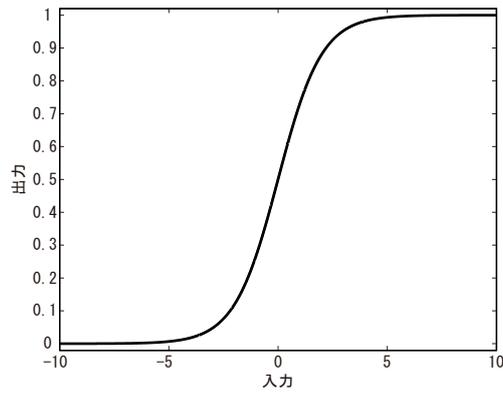


図 7: 標準シグモイド関数

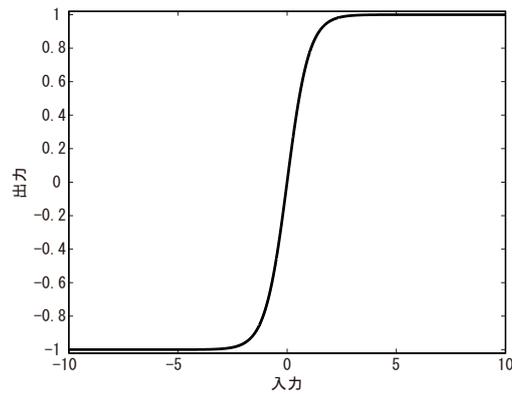


図 8: 双曲線正接関数

2.2.2 単純パーセプトロン

人工ニューロンをネットワーク状に繋げたモデルに単純パーセプトロンがある。このネットワークは F. Rosenblatt によって 1958 年に提案されたモデルであり、パターン認識の機能を有する [37]。

このモデルは図 9 に示すようにデータの入力を担う入力層と処理結果を出力する出力層の 2 層から構成されている。図中の x_k は k 番目の入力ニューロン (入力ニューロンは変数番号 k の入力データに対応する) を、 K は入力ニューロンの個数を示している。また、 u_i は i 番目の出力ニューロンへの入力を、 y_i は i 番目の出力ニューロンからの出力を、 I は出力ニューロンの個数を示しており、そして $w_{i,k}$ は k 番目の入力ニューロンから i 番目の出力ニューロンに接続されている重みを示している。

ニューロンには関数が設けられており情報処理が行われる旨は既に述べたが、入力層のニューロンではこれは行われず、データをそのまま出

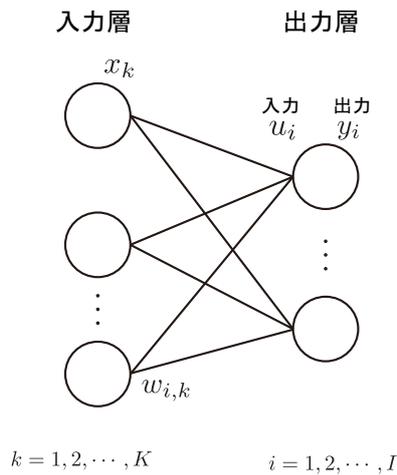


図 9: 単純パーセプトロン

力する。2.2.1 で述べた動作が表れてくるのは出力層からである。出力層のニューロンは入力ニューロンからの出力を受け取り、総和を計算してヘヴィサイド関数にて処理を行う。パーセプトロンは2層構造であるのでこれ以降の情報伝達は存在せず、よって出力ニューロンからの出力がこのネットワークからの最終出力となる。

こうして得られる最終出力であるが、この値が望む値かどうかというのは別問題である。例えばあるデータを入力して「1」という出力が欲しいとし、そしてネットワークにデータを入力した結果「0」が得られたとする。ニューラルネットワークは情報解析の手法であるので、設計上の期待する動作が求められるため、この場合ネットワークは期待通りの動作をしていないためデータ処理に役立たない。

2.2.3 学習

ネットワークに期待通りの動作をさせるためには「学習」という動作が必要になる。ニューラルネットワークは脳の情報処理を模したものであり、脳は学習を行い知識を得るものである。よってこのニューラルネットワークを模した人工ニューラルネットワークでも学習が行われる。学習には「入力データ」、「重み」、「教師データ」の三つが必要となる。教師データは図 9 中には示されていないが、これは「設計上の望む結果」を意味する値である。この教師データを学習に使用する学習は、教師あり

学習と呼ばれる。

ここで重みを $\mathbf{w} = \begin{bmatrix} w_{1,1} & \cdots & w_{i,k} & \cdots & w_{I,K} \end{bmatrix}^T$ 、入力データを $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_k & \cdots & x_K \end{bmatrix}^T$ とそれぞれベクトルで表すと、パーセプトロンの学習は [37]、

I. すべての重みを 0 に初期化する。

II. 重みと入力データを用いて出力層のニューロンへの入力を計算し、そして出力 y_i を得る。

III-a. y_i と教師データが一致した場合は次の入力パターンのデータに対して II の処理を行う。

III-b. 教師データが 0 である場合に y_i が 1 となった場合は重みを $\mathbf{w} = \mathbf{w} - \mathbf{x}$ より更新し、次の入力パターンのデータに対して II の処理を行う。

III-c. 教師データが 1 である場合に y_i が 0 となった場合は重みを $\mathbf{w} = \mathbf{w} + \mathbf{x}$ より更新し、次の入力パターンのデータに対して II の処理を行う。

という流れで行われる。ネットワークはこの流れを繰り返すことで、設計する値を出力するように学習される (重みが決定される) のである。

単純パーセプトロンはニューラルネットワーク黎明期のネットワークである。提案された当時は大きな話題となったが、「単純パーセプトロンは線形分離可能な問題しか処理できない」とその能力に限界が示されてしまい限界の克服が必要となってしまった。

2.2.4 多層パーセプトロン

単純パーセプトロンの限界を克服するため、多層パーセプトロンの「誤差逆伝播法 (Backpropagation, BP)」が Rumelhart らにより 1986 に提案された [38]。このネットワークは図 10 で示されるように、入力層と出力層の間に中間層が設けられた構造を一般的にしている。図中の x_k は $k (= 1, 2, \dots, K)$ 番目の入力ニューロン、 u_j と v_j はそれぞれ $j (= 1, 2, \dots, J)$

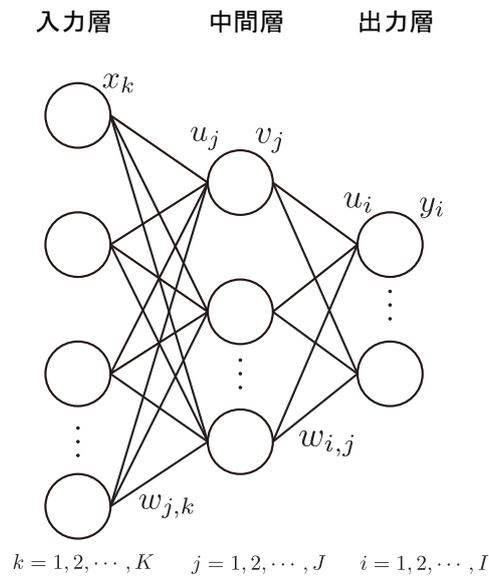


図 10: 多層パーセプトロン

番目の中間ニューロンへの入力と出力、 u_i と v_i はそれぞれ $i (= 1, 2, \dots, I)$ 番目の出力ニューロンへの入力と出力、 $w_{j,k}$ は k 番目の入力ニューロンから j 番目の中間ニューロンへの重み、 $w_{i,j}$ は j 番目の中間ニューロンから i 番目の出力ニューロンへの重みを示している。

このネットワークではニューロンで使用される関数が単純パーセプトロンとは異なる。単純パーセプトロンではヘヴィサイド関数を使用されていたが、誤差逆伝播法では中間層のニューロンに標準シグモイド関数や双曲線正接関数が一般的に用いられる。出力層のニューロンではネットワークの使用目的に応じて関数が選択される。数値予測用途では線形関数 ($f(x) = x$ で表される関数 (図 11))、分類用途では標準シグモイド関数や後述するソフトマックス関数を使用される場合が多い。

2.2.5 誤差逆伝播法

2.2.5.1 数値を予測する場合

単純パーセプトロンでは層が2層であり、重みの更新は入力層-出力層の間のみで行われたが、誤差逆伝播法では入力層-中間層、そして中間層-出力層の2層で行われる。また、誤差の測定尺度も単純パーセプトロンとは異なっており、誤差逆伝播法では誤差の測定に平均2乗誤差 (Mean

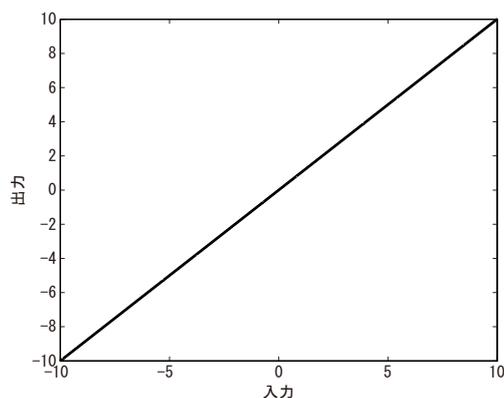


図 11: 線形関数

squared error, MSE) が使用される。図 10 で示されるネットワークの場合、入力データ $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_k]^T$ がネットワークに入力された際の出力 $y(\mathbf{x})$ に対する誤差 E は

$$E = \frac{1}{2} (y(\mathbf{x}) - t)^2$$

となる。ここで t は教師データである。誤差逆伝播法ではこの E を用いて入力層-中間層間の重みと中間層-出力層間の重みを更新する。

ネットワークの出力を得る計算は入力層から順に出力層までを伝播していくが、学習で重みを修正する際は出力層から逆に層を伝播していくので、誤差逆伝播法という名称がついている。ここで、誤差逆伝播法による学習の流れを述べる前に順伝播について述べる [36]。

ネットワークは各層間の重みがランダムに与えられた後、動作を開始する。入力層のニューロンからは単純パーセプトロンと同じく値が出力されてくるのみであるので、計算が行われるのは中間層の各ニューロンからである。 j 番目の中間ニューロンへの入力 u_j は、

$$u_j = \sum_{k=1}^K w_{j,k} x_k \quad (2-9)$$

により求められる。中間ニューロンへの入力は関数 $f(x)$ (ここでは標準シグモイド関数を使用するものとする) を経て中間ニューロンからの出力

となるので、 j 番目の中間ニューロンからの出力 v_j は、

$$v_j = f(u_j) \quad (2-10)$$

で求められる。

続いて計算は中間層-出力層間に移る。 i 番目の出力ニューロンへの入力 u_i は、

$$u_i = \sum_{j=1}^J w_{i,j} v_j$$

により求められる。そして、出力ニューロンへの入力は関数 $f(x)$ (ここでは線形関数を使用するものとする) を経て出力ニューロンからの出力となるので、 i 番目の出力ニューロンからの出力 y_i は、

$$y_i = f(u_i)$$

で求められる。ここで、出力ニューロンで使用される関数は入力がそのまま出力となる線形関数であることから、ネットワークからの出力 y_i (i 番目の出力ニューロンからの出力) は、

$$\begin{aligned} y_i &= u_i \\ &= \sum_{j=1}^J w_{i,j} v_j \end{aligned} \quad (2-11)$$

となる。

ネットワークからの出力が求まったら、逆伝播により重みを更新する。以降、更新の流れ [36] を述べる。はじめに出力層-中間層の重みの更新を考える。出力層からは最終的なネットワーク出力が出力されるので、まずはネットワーク出力と教師データとの誤差を求める。誤差は

$$E = \frac{1}{2} \sum_{i=1}^I (y_i - t_i)^2 \quad (2-12)$$

で示す平均2乗法を用いて計算されるとした場合、出力層-中間層間の重

み更新量 $\Delta w_{i,j}$ は式 (2-12) を $w_{i,j}$ で微分することで計算される。 $\Delta w_{i,j}$ は

$$\begin{aligned}\Delta w_{i,j} &= \frac{\partial E}{\partial w_{i,j}} \\ &= \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial w_{i,j}}\end{aligned}\quad (2-13)$$

であり、右辺第1項は

$$\frac{\partial E}{\partial y_i} = y_i - t_i \quad (2-14)$$

である。また右辺第2項は

$$\frac{\partial y_i}{\partial w_{i,j}} = v_j \quad (2-15)$$

より、

$$\begin{aligned}\Delta w_{i,j} &= \frac{\partial E}{\partial w_{i,j}} \\ &= (y_i - t_i)v_j\end{aligned}\quad (2-16)$$

と求められる。重みは $\Delta w_{i,j}$ を求めた後

$$w_{i,j} = w_{i,j} + \Delta w_{i,j} \quad (2-17)$$

により更新される。式 (2-17) の左辺の $w_{i,j}$ は更新後の重みを、右辺の $w_{i,j}$ は更新前の重みである。

続いて中間層-入力層間の重み更新量 $\Delta w_{j,k}$ であるが、これは

$$\begin{aligned}\Delta w_{j,k} &= \frac{\partial E}{\partial w_{j,k}} \\ &= \frac{\partial E}{\partial u_j} \frac{\partial u_j}{\partial w_{j,k}}\end{aligned}\quad (2-18)$$

より得られる。右辺の第1項は

$$\frac{\partial E}{\partial u_j} = \sum_{i=1}^I \frac{\partial E}{\partial u_i} \frac{\partial u_i}{\partial u_j} \quad (2-19)$$

であり、このうち右辺の第1項は $y_i = u_i$ より

$$\frac{\partial E}{\partial u_i} = u_i - t_i \quad (2-20)$$

となる。また右辺第2項は $u_i = \sum_{j=1}^J w_{i,j} f(u_j)$ より

$$\frac{\partial u_i}{\partial u_j} = w_{i,j} f'(u_j) \quad (2-21)$$

となる。さらに式 (2-18) の右辺第2項は

$$\frac{\partial u_j}{\partial w_{j,k}} = x_k \quad (2-22)$$

となるので

$$\begin{aligned} \Delta w_{j,k} &= \frac{\partial E}{\partial w_{j,k}} \\ &= \left(f'(u_j) \sum_{i=1}^I w_{i,j} (u_i - t_i) \right) x_k \end{aligned} \quad (2-23)$$

と求められる。なお、標準シグモイド関数の導関数は

$$f'(u_j) = x_k(1 - x_k) \quad (2-24)$$

である。重みは $\Delta w_{j,k}$ を求めた後

$$w_{j,k} = w_{j,k} + \Delta w_{j,k} \quad (2-25)$$

より更新される。式 (2-25) の左辺の $w_{j,k}$ は更新後の重みを右辺の $w_{j,k}$ は更新前の重みを意味する。以上のように順伝播と逆伝播を繰り返すことで、最終的にネットワークは目的とする値を出力するようになる。

以上が単純パーセプトロンを改良し構築された多層パーセプトロンの誤差逆伝播法の動作概要である。誤差逆伝播法は単純パーセプトロンの限界を克服し、線形分離不可能な問題も解くことが可能である。

2.2.5.2 クラスを予測する場合

2.2.5.1 では MSE を誤差測定に使用したが、これは数値予測を行うネットワークの一例である。ここではクラス分けを行うネットワークで一般的に使用される交差エントロピーについて述べる [36]。

数値予測を行うネットワークにおいて出力ニューロンの伝達関数は線形関数であったが、この関数を変更することにより、クラス分けを行う多層パーセプトロンが構築できる。多クラス分けを行うネットワークの場合、クラスの数だけ出力ニューロンを用意し、また、伝達関数にソフトマックス関数と呼ばれる関数が使用される場合が多い。ソフトマックス関数は x_i ($i = 1, 2, \dots, I$) を入力値とすると

$$f(x_i) = \frac{\exp(x_i)}{\sum_{l=1}^I \exp(x_l)} \quad (2-26)$$

で表される関数である。

ソフトマックス関数を使用した場合、 i 番目の出力ニューロンからの出力である y_i は、ニューロンへの入力 u_i より

$$y_i = f(u_i) \quad (2-27)$$

で求められる。このソフトマックス関数を経て出力されるネットワークからの出力値の総和は 1 となる性質があり、各出力ニューロンから出力される値 (クラス) は、入力データ \mathbf{x} が各クラスに属している確率と解釈することができ、各クラスを C_i ($= C_1, C_2, \dots, C_I$) で表すならば

$$p(C_i|\mathbf{x}) = y_i \quad (2-28)$$

と表すことができる。この多クラス分けを行うネットワークは i 番目から I までの各出力ニューロンから各クラスに属する確率が出力されることより、 s パターン目の入力に対するネットワークのターゲットベクトルは $\mathbf{t}^s = \begin{bmatrix} t_1^s & \dots & t_I^s \end{bmatrix}^T$ で表される。すると、入力 \mathbf{x}^s が入力された時の重み \mathbf{w} の尤度は

$$L(\mathbf{w}) = \prod_{s=1}^S p(\mathbf{t}^s | \mathbf{x}^s; \mathbf{w}) \quad (2-29)$$

となり、この尤度関数を最大化する \mathbf{w} を見つけ出すことがネットワークの目標となる。誤差関数はこの尤度関数の対数を取り、符号を反転した

$$E(\mathbf{w}) = - \sum_{s=1}^S \sum_{i=1}^I t_i^s \log y_k(\mathbf{x}^s; \mathbf{w}) \quad (2-30)$$

を使用する。符号を反転するのは最小化問題とするためである。この誤差関数は交差エントロピーと呼ばれる。

2.2.6 早期停止法

早期停止法 [36] は過学習を抑制する手法の一つである。ニューラルネットワークは学習を繰り返すことで与えられた入力データと教師データの関係进行学习し、出力と教師データとの誤差を小さくしていく。ところが、学習を幾度も繰り返して、与えられたデータについてあまりにも忠実に学習してしまうと、未知なる入力値が入力された場合に出力値が望む出力値と大きく乖離してしまう「過学習」という問題が発生する。

早期停止法はこの過学習を抑制するための手法であり、ネットワークの学習毎にネットワークに対して未知のデータを与えて出力値を監視し、誤差の発生状況に応じてネットワークの学習を停止させる方法である。早期停止法を採用する場合、誤差を監視するために監視用データが必要である。早期停止法は学習中のネットワークに監視用データを入力し、その誤差を監視、誤差が増加し始めたら過学習が生じていると判断して学習を停止させる。

2.2.7 深層学習

多層パーセプトロンは一般的に3層構造をしているがこれは3層で制限されているというのではなく、より多くの層を設けることが可能である。ところが、層数が増加し深い構造のネットワークとなると学習が

困難となる問題が生じてくる。この問題は勾配消失問題 [36] と呼ばれており、誤差を逆伝播させる際に重みの更新が困難になるという問題である。誤差逆伝播法は誤差を逆伝播させることにより重みを更新し学習を進めていくため、重みの更新が困難になることはすなわち、学習ができないことを意味する。

この勾配消失問題を極力回避するため、層数が多いネットワークでは事前学習と呼ばれる手法が採用される。事前学習とはニューラルネットワークの初期重みをランダムに与えるのではなく、予め学習した結果を用いるものである。事前学習の方法はさまざま存在しており、例えば自己符号化器を用いる方法や後述する自己組織化マップを用いる方法などである。

自己符号化器は図 12 に示す構造をしたネットワークである。特徴は入力と出力が同じものであることで、これは例えば 10 個の入力を 5 個の中間ニューロンを通して 10 個の入力を再現するというものである。すなわち、学習がうまく完了した自己符号化器において 5 個の中間ニューロンは、10 個の入力ニューロンの特徴を学習していることになる。この中間ニューロンの重みを層が深いネットワークの初期重みとして用いることで、勾配消失問題を回避して学習を進めることが可能となる。自己符号化器を多数用いることも可能であり、これは積層自己符号化器と呼ばれている。

2.2.8 自己組織化マップ

2.2.3 の単純パーセプトロンと 2.2.5 の誤差逆伝播法は学習に教師データが必要な教師あり学習のネットワークであったが、ニューラルネットワークには学習に教師データが不要な教師なし学習を行うネットワークも存在している。ここではこの教師なし学習について「自己組織化マップ」と呼ばれるネットワークを取り上げ述べていく。

自己組織化マップ [39] は 1982 年に T. Kohonen によって提案されたネットワークであり、図 13 に示すような構造をしている。自己組織化マップ

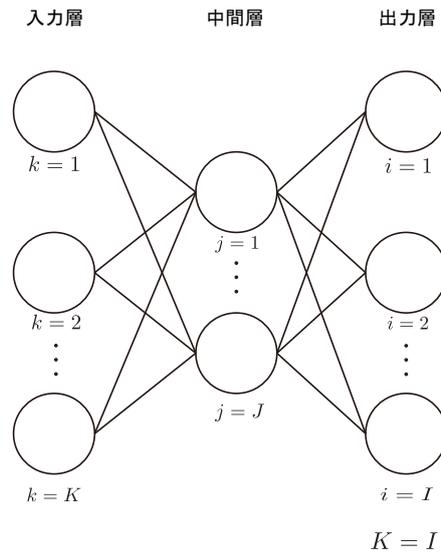


図 12: 自己符号化器

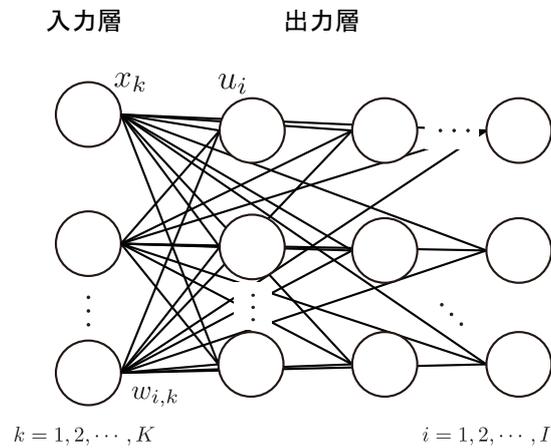


図 13: 自己組織化マップ

は入力ニューロンからなる入力層と、出力ニューロンからなる出力層の2層構造をしており、出力層の各出力ニューロンは入力層の各入力ニューロンと重みを介して接続されている。

自己組織化マップは出力ニューロンの重みを入力値に近づけるように更新を行う。これにより、ある特定の入力に反応する出力ニューロンが作成され、入力データの分類が結果として得られる。よって、自己組織化マップは主に分類器として使用されている。

自己組織化マップ (Self-organizing maps, SOM) の学習は入力データと最も強く反応する (関係が近い) 出力ニューロンを探すことから始まる [39]。 $k (= 1, 2, \dots, K)$ 番目の入力ニューロンを x_k 、 k 番目の入力ニュー

ロンから $i (= 1, 2, \dots, I)$ 番目の出力ニューロンへ接続されている重みを $w_{i,k}$ で表し、そして \mathbf{x} を入力データベクトル、 \mathbf{w}_j を j 番目出力ニューロンの重みベクトルとすると、出力ニューロンと入力ニューロンとの反応度合いは距離 ($\|\cdot\|$) を用いて (ここでは距離はユークリッド距離を用いる)

$$\|\mathbf{x} - \mathbf{w}_i\| = \sqrt{\sum_{k=1}^K (x_k - w_{i,k})^2} \quad (2-31)$$

より求められる。検索すべきは最も強く反応する出力ニューロンであるので、これは距離が最も小さい i 番目の出力ニューロンを探すことにより得られる。すなわち

$$c = \arg \min_i \{\|\mathbf{x} - \mathbf{w}_i\|\} \quad (2-32)$$

となる。 c は最も入力ニューロンと関係が近い出力ニューロン (Best-Matching Unit, BMU) を示す。

BMU の検索が完了すると、続いて重みの更新が行われる。ここで $t (= 1, 2, \dots, T)$ を時間とすると、重みの更新は

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + h_{c,i}(t)[\mathbf{x}(t) - \mathbf{w}_i(t)] \quad (2-33)$$

により行われる。ここで $h_{c,i}$ は近傍関数であり

$$h_{c,i} = e^{-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}} \quad (2-34)$$

で表される。またここで r_c は BMU の位置、 r_i は i 番目のニューロンの位置を表している。近傍関数を重み更新に使用することにより、BMU であるニューロンはより入力データに近づき、BMU 付近に位置するニューロンは入力データにある程度近づくという動作をすることになる。

以上の学習作業を繰り返すことにより、あるデータと近い関係にある出力ニューロン付近には、そのデータに近いデータが集まってくることになり、これによってデータの分類結果が得られる。

2.3 情報論的ニューラルネットワーク

2.3.1 情報理論

ここでは情報理論について述べる [40], [41]。情報理論は情報を数値化して表現する学問領域で、Claude Shannon が 1948 年に発表した「通信の数学的理論」に端を発している [41]。我々は「情報量が少ない広告」などと、情報の多さ・少なさについて日常的に口にしますが、そもそも日々のコミュニケーションや・広告、新聞といったものから発信される情報は重さや大きさといった物理的な性質は持っていない。ただし、「商品についての詳しい解説」、「簡易な解説」などと、我々は情報に量という概念を間違いなく感じている。情報理論はこの情報について、量や振る舞いなどを定義するものである。情報理論では確率 $P(E)$ を基に情報量 $I(E)$ を

$$\begin{aligned} I(E) &= \log_2 \frac{1}{P(E)} \\ &= -\log_2 P(E) \end{aligned} \quad (2-35)$$

と定義する (単位 (bit))。

情報理論の定義では情報は確率から算出される。そのため「ある情報はある人にとって情報量が多いが、ある人にとっては少ない」などというブレは生じない。反面このことは、我々が日常口にしている情報量とは概念が異なるものであることに注意が必要である。しかし情報の価値という観点から見ると情報量は、珍しい事柄についての情報では情報量は高い値を、珍しくない事柄については情報量は低い値をとるため、値の大きい方がより情報を持っていると直感的に理解できる。例えばコイントスを行う場合、コインが表か裏になる確率は $1/2$ であるので情報量は 1bit ということになるが、サイコロを振る場合、各目が出る確率は $1/6$ の確率で起こるため情報量は 2.585bit となり、これは $1/2$ の確率で起こる事柄の情報量 1bit よりも大きい。

ここで情報量の平均を計算することについて考える。例えば延々と何

かが繰り返される状態 $S (= s_1, s_2, \dots, s_i, \dots, s_I)$ の平均情報量 $H(S)$ は

$$\begin{aligned} H(S) &= \sum_{i=1}^I P(s_i) \log_2 \frac{1}{P(s_i)} \\ &= - \sum_{i=1}^I P(s_i) \log_2 P(s_i) \end{aligned} \quad (2-36)$$

より算出される。ここで s_i は出現する事象である。この平均情報量はエントロピーと呼ばれる。情報量は珍しい事柄が生じると大きい値を持つものであるが、これは逆に考えると、不確かさが大きいとも解釈できる。つまりエントロピーは不確かさの平均ともとることができる。

続いて、ある情報を得たときにその情報からどのくらいの情報を得たのかということについて考えてみる。これは例えば外の天気は全く分からない状態におかれている人物が、天気予報から「本日は晴れる見込み」と予報された情報を受け取った場合に、天気予報からどのくらいの情報量を得たのかを求めることである。すなわち、不確かな実際の天気が、天気予報という予報情報を得ることにより、どれだけ減少するかと解釈できる。情報を得ることで不確かさが減るという状況の情報量を相互情報量という。実際の天気を X 、天気予報を Y 、実際の天気のエントロピーを $H(X)$ 、天気予報のエントロピーを $H(Y)$ 、天気予報を知ったときの実際の天気のエントロピーを $H(X|Y)$ 、実際の天気を知ったときの天気予報のエントロピーを $H(Y|X)$ で表すと相互情報量 $I(X; Y)$ は図 14 に示す関係、すなわち

$$I(X; Y) = H(X) - H(X|Y) \quad (2-37)$$

により求められる。

2.3.2 反復情報量最大化法

反復情報最大化法 [42] は自己組織化マップを情報理論を用いて改良したものである。自己組織化マップを用いると入力データの分類結果を得ることができるが、分類結果に影響を与えた入力変数までを把握するこ

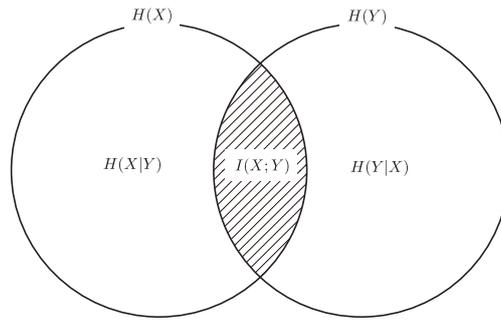


図 14: 相互情報量

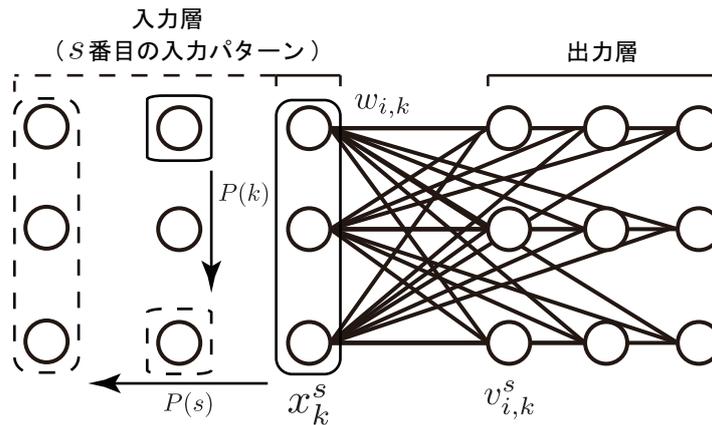


図 15: 反復情報最大化法概要図 (文献 [86] より引用)

とは困難である。しかし、反復情報最大化法では情報量に着目することで入力データ中の重要な変数に関する事柄を得ることができる。またこれは、パラメータを調整することにより、得られる情報を調整することも可能である。このことはデータ解析開始時の、まだデータについて何の前知識も持っていない場合に有効である。はじめに最も重要な変数のみを求め、続いてパラメータを調整し第二、第三の重要変数へと解析・解釈対象を拡大することでデータ構造を把握して解析することができるからである。

図 15 に反復情報最大化法の概要を示す。構造自体に自己組織化マップと差異はないが、動作に違いがあり情報量に着目して学習を行うことが可能となっている。

反復情報最大化法は出力層のニューロンからの出力を

$$v_{i,k}^s = \exp\left(-\sum_{k=1}^K \frac{(x_k^s - w_{i,k})^2}{2\sigma_{k,l}^2}\right) \quad (2-38)$$

と定義して学習を行う。ここで x_k^s を $s (= 1, 2, \dots, S)$ 番目の入力パターン $k (= 1, 2, \dots, K)$ 番目からの入力ニューロン、 $w_{i,k}$ を $i (= 1, 2, \dots, I)$ 番目の出力ニューロンへ k 番目の入力ニューロンから接続されている重み、 $\sigma_{k,l}^2$ は獲得情報量を調整するパラメータである。また l は、 k と同じく入力ニューロンのインデックスを表しているが、 l で表された場合は l 番目の入力ニューロンに着目している状態を指す。

この「ある入力ニューロンに着目すること」を反復情報最大化法では「エンハンスメント」と称している。エンハンスメントは式 (2-38) のパラメータ σ によってコントロールされる。パラメータ σ は

$$\sigma_{k,l} = \begin{cases} \frac{1}{\beta}, & k = l \text{ (エンハンスメント)} \\ \beta, & \text{otherwise} \end{cases} \quad (2-39)$$

の値を取る。なお、パラメータ β は ($\beta > 0$) の値である。

求められた $v_{i,k}^s$ は

$$p(i|s; k) = \frac{v_{i,k}^s}{\sum_{m=1}^I v_{m,k}^s} \quad (2-40)$$

により標準化され出力ニューロンの発火確率が計算される。続いて出力ニューロンの発火確率を用いて、

$$MI(k) = \sum_{s=1}^S \sum_{i=1}^I p(s) p(i|s; k) \log \frac{p(j|s; k)}{p(j; k)} \quad (2-41)$$

より相互情報量 $MI(k)$ が求められる。ここで $p(i; k)$ は

$$p(i; k) = \frac{1}{S} \sum_{s=1}^S p(i|s; k) \quad (2-42)$$

である。そして最後に入力ニューロンの発火確率 $p(k)$ が

$$p(k) = \frac{MI(k)}{\sum_{l=1}^I MI(l)} \quad (2-43)$$

より求められる。この求められた $p(k)$ は学習時に、

$$\| \mathbf{x} - \mathbf{w}_i \| = \sqrt{\sum_{k=1}^K p(k) (x_k - w_{i,k})^2} \quad (2-44)$$

で用いられる。これは自己組織化マップで示した式 (2-31) に入力ニューロンの発火確率を組み込んだものであり、すなわちニューロンの状態を加味して学習を行うことになる。

また、学習終了後にこのエンハンスメントを確認することで反復情報最大化法が学習時 (データ分類時) に活用した変数について知ることができる。この場合、最も発火確率が高い入力ニューロン (変数) が最も重要な変数と解釈する。

2.3.3 潜在学習法

潜在学習法 [43] は反復情報最大化法を発展させたものである。反復情報最大化法はデータを分類し、どの変数が重要であるかを調べる機能を有しているが、データの予測機能は有していない。よって、クラスタリング問題でしか活用することができないという問題を抱えていた。

潜在学習法は図 16 に示すように二つの学習段階から構成されている。一つめの段階は自己組織化マップがベースになっているネットワーク構造を、二つめの段階は多層パーセプトロン (誤差逆伝播法 (BP)) がベースになっているネットワーク構造をしている。学習 1 段階目は潜在性 (定義は後述する) の高いニューロンを探し出すと共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。学習 2 段階目は学習 1 段階目で得られた知識 (学習結果) を用いてターゲットデータについて予測する。

学習 1 段階目で計算される潜在性は入力ニューロンと出力ニューロンの

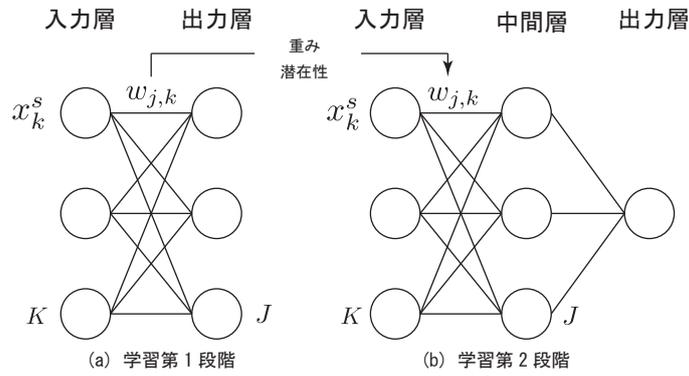


図 16: 潜在学習法概要図 (文献 [43] より引用)

重みの分散を基に算出される。 k 番目の入力ニューロンの分散 V_k は

$$V_k = \frac{1}{J-1} \sum_{j=1}^J (w_{j,k} - w_k)^2 \quad (2-45)$$

で定義する。ここで $w_{j,k}$ を k 番目の入力ニューロンから j 番目の出力層のニューロンへの重み、 J を出力層のニューロンの個数を意味している。また 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

$$w_k = \frac{1}{J} \sum_{j=1}^J w_{j,k} \quad (2-46)$$

から求められる。そして k 番目の入力ニューロンの潜在性を意味する ϕ_k^r は

$$\phi_k^r = \left(\frac{V_k}{\max_{k=1, \dots, K} V_k} \right)^r \quad (2-47)$$

より求められる。分母の V_k は k が 1 から入力ニューロンの個数である K までの時の分散の最大値である。また r は係数であり潜在性を強調する役割を持っている。よって ϕ_k^r はパラメータ r の時の k 番目の入力ニューロンの潜在性を意味している。

潜在性は式 (2-45) より k 番目の入力ニューロンの分散を求め、それを用いて式 (2-47) よりパラメータ r の際の k 番目の入力ニューロンの潜在性が求められるわけであるが、パラメータ r が 1 であるとする k 番目ニューロンの潜在性とは、 K 個中で最も分散が高いニューロンからみてどのくらい k 番目ニューロンの分散はあるのかということの意味してい

る。これは分散が大きいニューロンが重要な役割を果たしているという考えに基づいており、分散を重要とするのはニューラルネットワークに情報理論を持ち込んだパイオニアである Linsker [33] に沿うものである。

このようにして求められた潜在性は自己組織化マップに組み込まれ、第1段階の学習が行われる。潜在学習法の学習1段階目は距離を算出する際に

$$\| \mathbf{x}^s - \mathbf{w}_j \| = \sqrt{\sum_{k=1}^K \phi_k^r (x_k^s - w_{j,k})^2} \quad (2-48)$$

を用いて距離を求める。この後の学習方法は通常の自己組織化マップと同様である。距離を求める際に ϕ_k^r を用いているが、潜在学習法では潜在性が高いニューロンを重要視するように学習が行われ、この点で通常の自己組織化マップと異なっている。

自己組織化マップは入力ニューロンと重みの距離を求め学習が行われる。通常の自己組織化マップでは距離のみに着目して BMU を探索するが、潜在学習法ではこの処理に潜在性を加味して BMU を探索する。潜在性は最大値が1であるので、最も潜在性の高いニューロンが距離算出に使用される場合、求められた距離そのものが BMU 探索に用いられる。しかし、潜在性が低いニューロンが距離算出に使用される場合は求められた距離が減じられて BMU 探索に使用されることになる。潜在性により重みづけられた距離を用いることにより潜在性に着目した学習を期待している。また学習1段階目に自己組織化マップを使用することであるが、これは学習2段階目に渡す知識(重み)をわかりやすいものとするを期待してのことである。自己組織化マップの学習結果は入力データから特徴を抽出したものである。よって、学習2段階目にとってはこれから解析するデータについての情報であり、これを作り出す目的で学習1段階目に自己組織化マップを使用する。

以上の流れで学習1段階目の学習は終了である。潜在学習法は続いて2段階目の学習へ移行する。学習2段階目はターゲットデータを予測する働きをし、これは多層パーセプトロンの誤差逆伝播法を基に学習を行う。通常の誤差逆伝播法との違いは、入力層のニューロンと隠れ層のニューロ

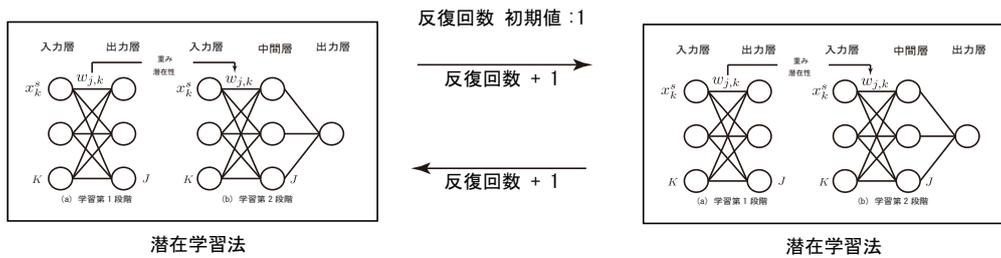


図 17: 反復潜在学習法の概要図

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 [44], [45]、乱数の中から最適な重みを探す [46], [47]、他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 [48]、などさまざまな試みが行われている。潜在学習法では学習1段階目で得られた重みを使用するわけであるが、これはそのまま学習2段階目の初期重みとするのではなく、重みに潜在性を乗算した値 (知識) を使用する。これは潜在学習法の学習2段階目の初期重みに学習1段階目で得られた知識を用いるという意味があり、このことにより学習1段階目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。

2.3.4 反復潜在学習法

反復潜在学習法 [49], [50] は潜在学習法を改良したものである。潜在学習法には獲得する情報量が十分に増加しないという問題が存在する。これは潜在学習法は獲得した情報の活用が十分でないためである。そこで反復学習により情報を活用する反復潜在学習法が考案された (反復潜在学習法の概要を図 17 に示す)。また、反復潜在学習法では潜在学習法と異なり中間ニューロンの潜在性を算出する。反復潜在学習法は図 17 に示すように、潜在学習法が繰り返されることにより情報量がより蓄えられるイメージである。

j 番目の中間ニューロンの潜在性 ϕ_j は j 番目の中間ニューロンの分散

v_j を基に求められる。分散は

$$v_j = \frac{1}{K-1} \sum_{k=1}^K (w_{j,k} - w_j)^2 \quad (2-49)$$

より計算される。ここで k は k 番目の入力ニューロン、 K は入力ニューロンの個数、 $w_{j,k}$ は j 番目の中間ニューロンへ接続されている k 番目の入力ニューロンからの重みである。また、 w_j は $w_{j,k}$ の平均であり、

$$w_j = \frac{1}{K} \sum_{k=1}^K w_{j,k} \quad (2-50)$$

より算出される。そして分散は

$$\phi_j = \left(\frac{v_j}{v_{max}} \right)^r \quad (2-51)$$

より標準化され、中間ニューロンの潜在性が得られる。式 (2-51) にある r は1 繰り返しあたりで獲得する情報量を制御するパラメータ ($0 \leq 1$) である。

反復潜在学習法は前述した潜在性を用いて潜在学習を行い、そしてさらにネットワークを繰り返し学習させることにより情報量を高めていく。この繰り返しとは学習が完了したネットワークをさらに学習させる反復学習のことで、またこの反復の際、反復回数1 回前時点の重みと潜在性を掛け合わせた値を新しい反復回数時点の重みとして使用することにより、反復ごとに情報量を高めていく。

3 情報論的ニューラルネットワークの検証

本章では本論文で提案する情報論的ニューラルネットワークのなかより、本論文の主提案手法となる潜在学習法を取り上げ、潜在性に着目することによる汎化能力への影響、抽出される重要変数の妥当性について、また多重共線性発生データ解析時の解析結果について検証を行う。また、潜在学習法と他の手法との違いについても述べる。潜在学習法の各種検証には人工的に作成したデータを用いた。ニューラルネットワーク手法による解析として、誤差逆伝播法 (BP)、重み減衰を用いた誤差逆伝播法 (WD)、潜在性を使用しない潜在学習法 (SBP)、潜在学習法 (PL)、ロジスティック回帰分析による解析として、変数削減を使用したロジスティック回帰分析 (Logi Reg)、変数削減を使用しないロジスティック回帰分析 (Logi Reg ALL) の汎化能力を比較した結果、ロジスティック回帰分析よりもニューラルネットワーク手法による解析の方がモデル汎化能力が高く、また、ニューラルネットワーク手法の中では潜在学習法が汎化能力が高い結果を得たため、潜在学習法の有効性が確認できた。また抽出結果の妥当性検証では、人工データに設けられた正解となる重要変数を潜在学習法により抽出することに成功した。さらに、潜在学習法に組み込んだロジックの動作を確認したところ、期待通りの動作をしていることの確認がとれた。くわえて、潜在学習法は動作ロジックが他の重要変数抽出能力を持った手法よりも平易であり、第三者に対する説明性・他手法への拡張性といった観点でも強みを持った手法であることが確認できた。

3.1 解析データと解析手法

3.1.1 解析データ

本章では潜在学習法について、潜在性に着目する効果の検証と抽出された重要変数の妥当性の検証を行う。この検証のため2変数からなるサンプル数1000の2クラス分けを目的とするデータを人工的に作成した。変数1 (X1) は目的変数が「0」である場合に1から2までの数値をラン

表 1: 人工データ (一部表示)

X1	X2	Y	Flg
1.1	3.2	0	0
1.6	6.6	0	0
1.7	4.5	0	0
1.8	6.1	0	0
1.3	4.2	0	0
～略～			
2.9	3.1	1	1
8.2	5	1	0
8.6	3.4	1	0
8.5	4.6	1	0
3.4	3.1	1	1

ダムでとり、また「1」である場合には8から9までの値をランダムでとる。変数2 (X2) は3から7までの値をランダムにとるもので、こちらの数値は目的変数に影響を与えないものである。このことは、変数1が1から2の範囲であるか、または8から9までの範囲であるかを確認することにより、目的変数の値が判明するというものであり、すなわち変数1が重要な変数であるというデータ設計である。ただしデータの完全分離を避けるため、ランダムにノイズを加えている。このノイズは目的変数の値が「0」の場合はランダムに6から9までの値をとり、「1」の場合は1から4までの値をとる (データの具体例は表1に示す。この表中の「Flg」が「1」である場合にノイズが混入されている。Flgは約20%の確率で1となる。また図18に散布図を示す)。

3.1.2 解析手法

本章ではこの人工データを誤差伝播法 (BP)、重み減衰を用いた誤差逆伝播法 (WD)、誤差逆伝播法の事前学習に自己組織化マップを用いた手法 (SBP)、潜在学習法 (PL)、変数削減を使用したロジスティック回帰分析 (Logi Reg)、全変数を用いたロジスティック回帰分析 (Logi Reg ALL) を用いて解析を行う。

重み減衰を用いた誤差逆伝播法は誤差逆伝播法の学習の際に重みの更

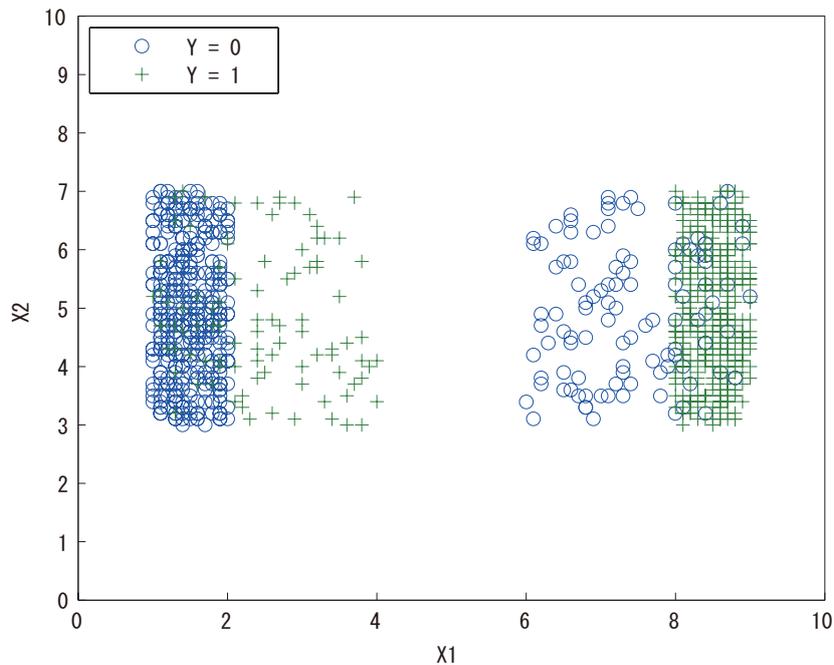


図 18: 人工データの散布図

新を制御する手法である。重み減衰は誤差関数 $E(\mathbf{w})$ に重みの二乗和を加え (L2 正則化と呼ばれている) $E(\mathbf{w}) + \frac{\lambda}{2}\|\mathbf{w}\|^2$ より実施され (λ は正則化の効き具合を調整するパラメータである)、重み更新の際に重みが減衰させられる [36]。重み減衰を用いることでネットワークは小さい重みを持つようになるため、正則化は過学習を抑制する目的で使用される手法であるが、学習終了時点まで残った重みと接続されている入力ニューロンは重要な入力であると捉えて本章では重要変数解釈目的で使用する。一部のニューロンだけに重みが存在する場合はブラックボックスであるニューラルネットワークの解釈も容易になるため、この効果による解釈を試みる。

誤差逆伝播法の事前学習に自己組織化マップ (SOM) を用いた手法 (SBP) であるが、これは、誤差逆伝播法の前学習として自己組織化マップを用いる (誤差逆伝播法の初期重みを自己組織化マップによって得る) 手法であると共に、潜在性を用いない潜在学習法と等価な手法である (2.3.3 で述べた式 (2-48) および学習第 2 段階の初期重み設定から潜在性に関する部分を除いた手法となる)。この手法を解析に使用することにより、ネットワークの学習に潜在性を使用することの効果について検証を行う。

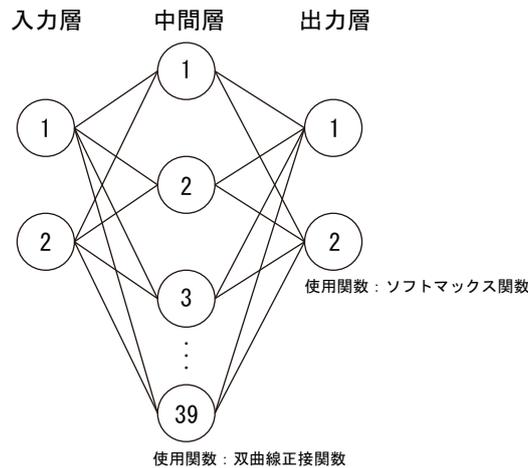


図 19: ネットワーク構成

3.1.3 使用したニューラルネットワークの構造

使用したネットワークの構成 (BP・WD・SBP の BP 部・PL の学習第 2 段階) は図 19 に示すものである。このネットワークは中間ニューロンが 39 個、出力ニューロンが 2 個から構成されたネットワークであり、中間ニューロンの個数は SBP と PL で使用する自己組織化マップのニューロン数より決定し、また出力ニューロンの個数は出力を 2bit で表現している (0 を 10、1 を 01) ため 2 個となっている。

自己組織化マップのニューロン数であるが、これは n をデータのサンプル数とし $5\sqrt{n} \times 0.25$ より [51], [52], [53] 決定された。自己組織化マップの出力ニューロン個数決定には理論的なルールが存在しないため [54]、本論文では Vesanto らが用いた方法 [51] を用いた。また、出力ニューロンの初期値はランダムに設定するのではなく「線形な初期化」 [39] を用いて設定した。これは入力データの広がり (主成分分析の第一主成分と第二主成分) から初期値を設定するもので、ある程度データに沿った初期値が設定されるため学習が容易になる。また初期値がランダム値ではなくデータから作り出されるため、学習毎に初期値が異なり学習結果が影響されるということがなくなる。ここで多次元のデータを 2 次元のマップに次元を削減していることについて 2 次元が最適なのかどうか指摘を受けるかもしれない。このことであるが、主成分分析において第一主成分は最も多くの情報を持った成分であり、第二主成分はそれに続く成分

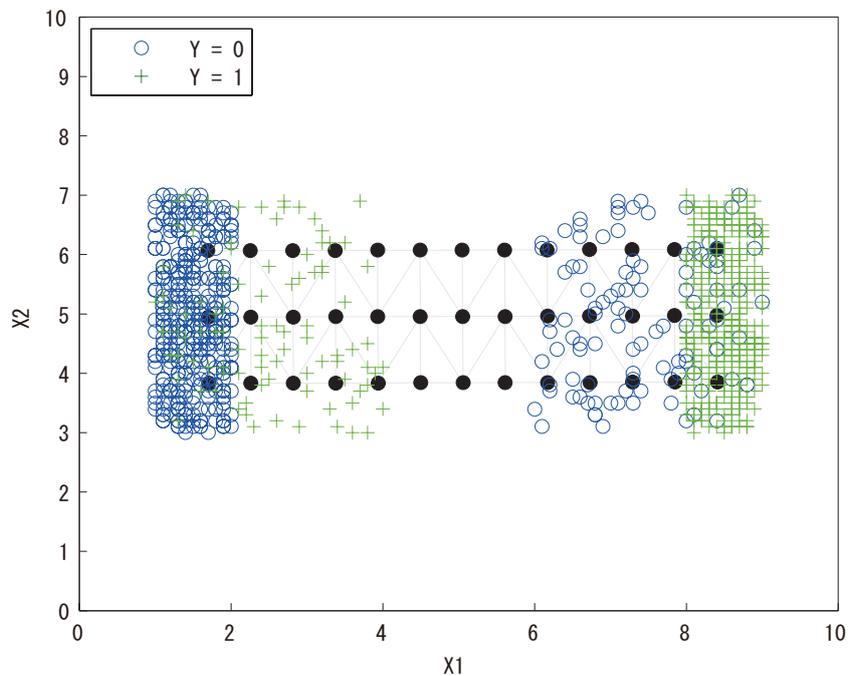


図 20: 人工データの散布図と自己組織化マップの出力ニューロン配置との関係 (学習前)

である。それ以降の主成分は算出はされるものの保持している情報は第一主成分、第二主成分ほどではなく小さいものである。主成分分析を実施する場合は累積寄与率が満足するまで任意の主成分までを使用するが、自己組織化マップの初期値設定ではデータの最大成分とそれに続く成分、すなわち、第一主成分と第二主成分でデータの大部分をカバーできていると考えており、2次元で十分であると考えている。図 20 は表 1 に示す学習データと当解析で使用する自己組織化マップの出力ニューロンについて示したものである (黒色で示す丸がニューロンを示している)。この図よりニューロンの配置 (初期値) がデータに沿っていることが確認できる。学習後の出力ニューロンについては後述する。

なお、潜在学習法で使用する潜在性であるが、潜在性算出に使用する分散はデータの分散ではなく出力ニューロンの重みの分散である。潜在学習法は学習開始時に潜在性を求めるため、使用する分散は出力ニューロンの初期値である。よって、潜在性は初期値に影響を受けることになる。線形な初期化により初期値はデータに沿って設定されているが、入力データに標準化を適用させる場合には標準化手法に注意を要する。平

均を 0、標準偏差を 1 とするような標準化手法を用いる場合、潜在性の算出結果が実データに沿った値にならない場合があるためである。なお本章の検証では初期化は実施していない (次章以降の解析事例では標準化を実施しているが、標準化手法については各章で適宜述べる)。

3.2 検証結果

3.2.1 汎化能力検証

ここでは人工データの解析結果より潜在学習法の汎化能力について確認を行う。表 1 に示すデータを誤差伝播法 (BP)、重み減衰を用いた誤差逆伝播法 (WD)、潜在性を用いない潜在学習法 (SBP)、潜在学習法 (PL)、変数削減を使用したロジスティック回帰分析 (Logi Reg) (ここでは各係数に対するカイ二乗検定の p 値を基準 (基準値 : 0.05) に有意でない変数を除いて変数選択を行った)、ロジスティック回帰分析 (Logi Reg ALL) にてそれぞれ解析した結果を図 21 に示す。なお、解析データは 1000 件の 70%にあたる 700 サンプルをモデル学習用、15%にあたる 150 サンプルを過学習を抑制するための早期停止用、15%にあたる 150 サンプルをモデル汎化能力測定用として用いた (WD はそれ自身が過学習を抑制するものであり、WD の場合は比率をモデル学習用が 85%、汎化能力検証用を 15%とした)。さらにこの分割比率は固定とし、各用途 (学習用、早期停止用、汎化能力検証用) にて使用されるデータサンプルをランダムに変更して 10 パターンのデータを作成した。これは良い結果をもたらすデータサンプルが偶然に使用されることを防止するためである。よって結果は 10 個のモデルから得られた平均値を示している。図 21 に示す結果より、人工データ解析モデルの汎化能力は高いものから、PL・SBP・BP・WD、そして両ロジスティック回帰分析 (Logi Reg ALL と Logi Reg) の順であることが確認できる。

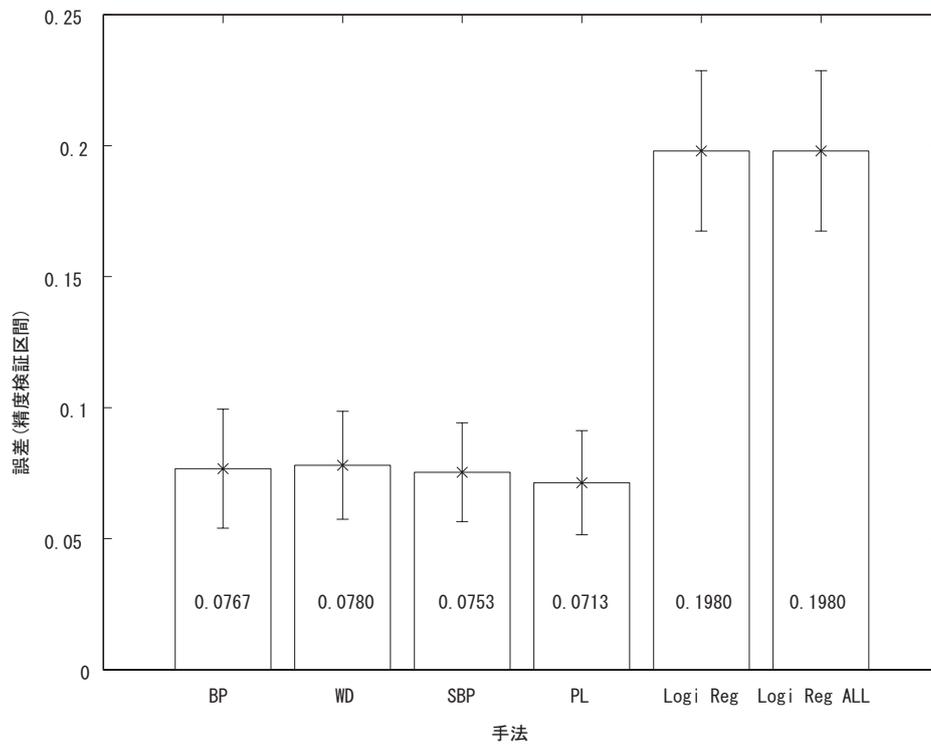


図 21: モデル汎化能力

3.2.2 抽出重要変数の妥当性検証

続いて重要変数の抽出について確認を行う。図 22 は各変数の重要性解釈のために必要な情報を示したものである。図 22 は入力層-中間層間の重み (図 22 (a) から (c))、潜在性 (図 22 (d))、ロジスティック回帰分析の係数 (図 22 (e) と (f)) を示している。図 22 (a) から (c) の各グラフの長方形が重みを意味しており、図形の大小で重みの大きさを、色で重みの符号 (黒: プラス、白: マイナス) を示している。また、X 軸は入力ニューロンの番号を、Y 軸は中間ニューロンの番号を表示している。また、これら図 22 の各図は 10 個のモデル中で最も良い汎化能力を示した際の結果を示している (誤差逆伝播法: 0.0533、重み減衰を用いた誤差逆伝播法: 0.0533 (パラメータ $\lambda = 0.9$)、潜在性を用いない潜在学習法: 0.0467、潜在学習法: 0.0467 (パラメータ $r = 1.8$)、ロジスティック回帰分析 (変数選択あり): 0.2333、ロジスティック回帰分析 (変数選択なし): 0.2333)。なお、図 22 (d) に重みではなく潜在性を記載しているのは潜在学習法では変数解釈に潜在性を用いるためである。

また、潜在性を用いない潜在学習法と潜在学習法は、自己組織化マッ

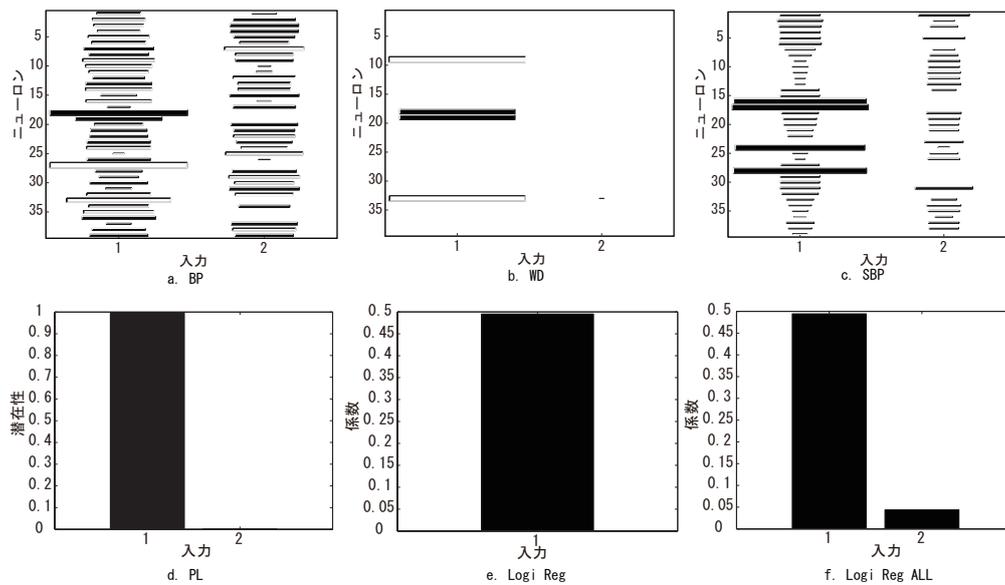


図 22: 重み (入力層-中間層)

プを用いて事前学習を行うため、自己組織化マップの学習結果も図 23 に示す。この図中の左側にある図は U-Matrix という方法により学習後の出力ニューロンの状態を示したものであり、右側の図は U-Matrix に対応する出力ニューロンの配置 (Map grid) を示したものである。Map grid に存在する黒色は出力ニューロンを、そして U-Matrix に存在している六角形の図形も出力ニューロンを示している。また、U-Matrix には Map grid 中ではニューロンが存在しない場所にも六角形が存在しているが、これは隣り合うニューロンとの距離を示すために設けられたものである。ここで用いた自己組織化マップはニューロンの近傍を六角形としているため、各ニューロンから見て 6 方位が近傍を求める対象となる。自己組織化マップはデータをクラスタリングする働きを持っているが、この U-Matrix によりクラスタについて確認をすることができる。図中の寒色は隣り合うニューロンと距離が近いことを、暖色は遠いことを示している。すなわち、暖色がクラスタを分断する境界線と捉えることができる。

自己組織化マップは「近い関係にあるものを近くに集める」という動作原理で作動する。そのため、あるデータに反応したニューロンの周りには反応したデータと近い関係にあるニューロンが配置され、クラスタが形成されることになる。つまり基本的には近い特徴を持つニューロン

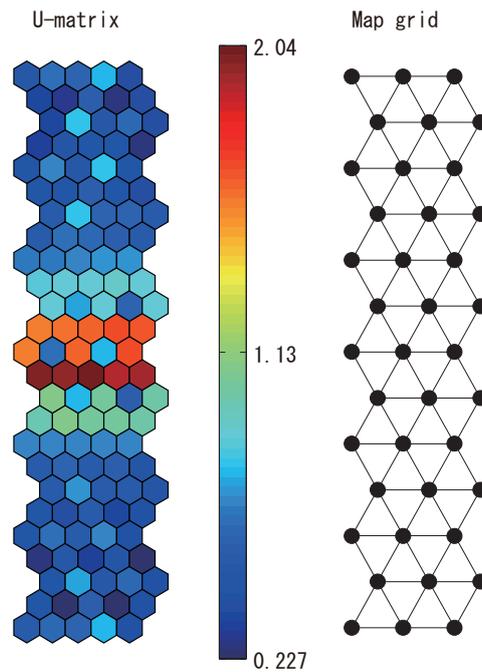


図 23: SBP における自己組織化マップの学習結果

が隣に存在していると考えられる。ところが、必ずしも隣り合っているからといって近い特徴ではなく、別の特徴を持つという場合も存在する。この、隣と自分が近い特徴を持っているかどうかを調べるために近傍について調べ (距離)、クラスタリング状態を可視化する方法を U-Matrix [55] という。参考までに図 24 に学習後の自己組織化マップの出力ニューロンとデータとの関係を示す。図からは散布図中心付近にある、データが分布していない箇所に三つのニューロンが存在していること、そしてそれを境としてクラス 0 のデータとクラス 1 のデータに分かれていること (クラスタが形成されていること) が確認できる。この状態を U-Matrix で表示したものが図 23 である (図 24 の出力ニューロンと図 23 の U-Matrix が 90 度回転しているのは表示上の都合である)。当解析で使用しているデータは散布図から特徴が捉えられるものであり、図 23 を用いなくてもデータの分類結果 (クラスタ) は確認できる。しかし、より込み入ったデータを解析する場合、図 24 のようにデータと出力ニューロンについてプロットした図を確認するよりも U-Matrix の方が学習結果の確認を行いやすく、U-Matrix は自己組織化マップを用いた解析で多用される図である。

図 23 より、データのクラスタはマップの中心で上下に分割されている

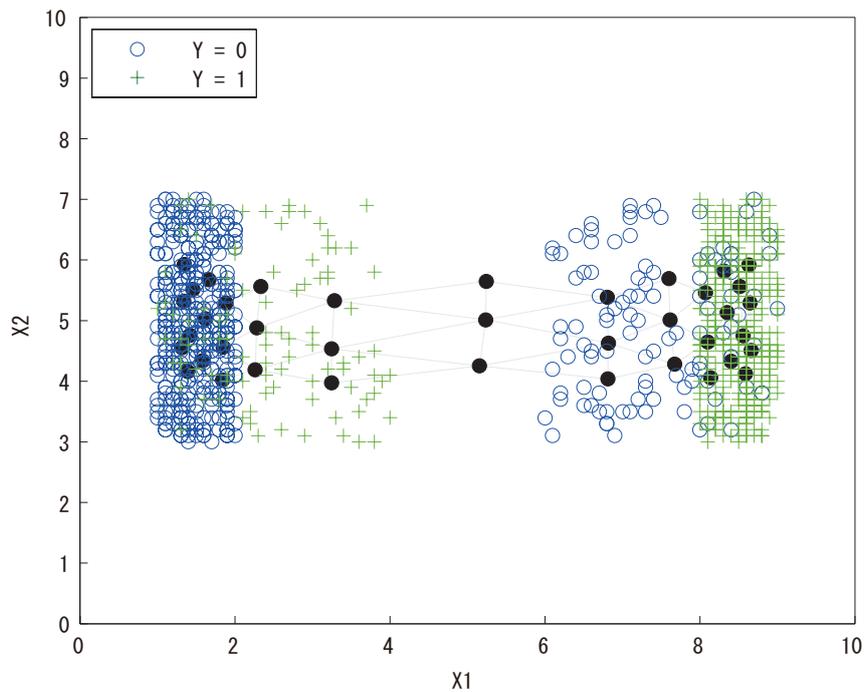


図 24: 人工データの散布図と自己組織化マップの出力ニューロン配置との関係 (学習後)

ことがわかる。潜在性を用いない潜在学習法はこの学習結果を基にクラスを予測させるため、事前学習を行わない誤差逆伝播法 (初期重みがランダムに設定される) と比べて分類精度が高い結果となっている (図 21)。

図 25 は潜在学習法で使用された自己組織化マップの結果である。潜在学習法は式 (2-47) に示す式により潜在性を求め、潜在学習法の学習第 1 段階 (自己組織化マップ) の学習の際に潜在性に着目をして学習を行う (式 (2-48))。この学習の初期状態には自己組織化マップの学習結果 (図 23) が用いられ、この状態から潜在性に着目した潜在学習法の学習第 1 段階の学習がはじまる。またこの際パラメータ r により (式 (2-47))、潜在性が強調される (当解析では r は 1 から 2 まで 0.1 ずつ増加 (すなわち、11 回 r を変更) させた)。このパラメータ r の変更は連続的 (R を現在のパラメータ r 変更回数とすると R 回目の学習の際は $R - 1$ 回目の学習結果が初期状態となる) に行われる。

図 25 は図 23 の学習結果を潜在学習法の初期状態とし、式 (2-47) に示すパラメータ r を増加させながら学習を行った際の図である。最高精度を得たのは図 22 の際、すなわち、 $r = 1.8$ の際であるので、図 25 はこ

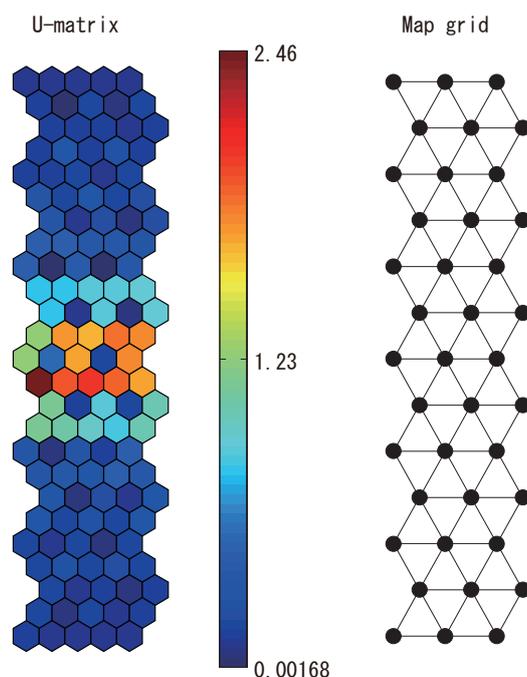


図 25: PL における自己組織化マップの学習結果

の時の結果を U-Matrix で示したものである。 $r = 1.8$ の際の潜在性は X1 が 1、X2 が 0 のため、学習には二つの変数のうち X1 のみが用いられていることになる。図 25 の U-Matrix は、潜在性を用いない潜在学習法の U-Matrix (図 23) と比較して、寒色がよりはっきりと表れた結果となった。すなわち分類がよりはっきりした結果が得られたことになる。この違いは潜在性によるものであり、潜在性を使用することの有効性が示されたと言える。クラスタの境界については変わりなく、出力ニューロンの中心付近であった。当解析では X1 のみが学習に活かされているという結果を得たため、参考までに X1 のみを用いて自己組織化マップの学習を行ったところ、図 26 に示す結果を得た。この図より U-Matrix 下部と上部は濃い寒色となっていることが確認でき、図 25 の傾向と違いはなかった。そのため、X1 のみを使用した結果として図 25 に示す結果が得られることの妥当性の確認ができた。

続いて入力-中間層の重みの確認を行う。図 22 の各結果より

誤差逆伝播法

BP の重みは各重みでそれぞれの重みをもっており変数の重要性について傾向などを確認することは困難であった。

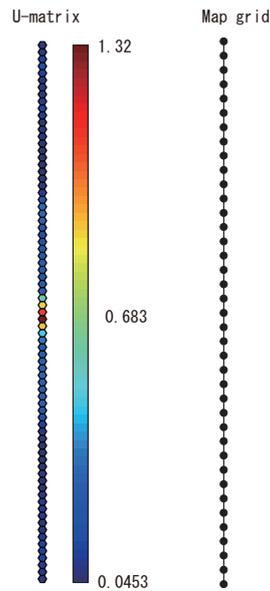


図 26: X1 のみを用いた際の自己組織化マップの学習結果

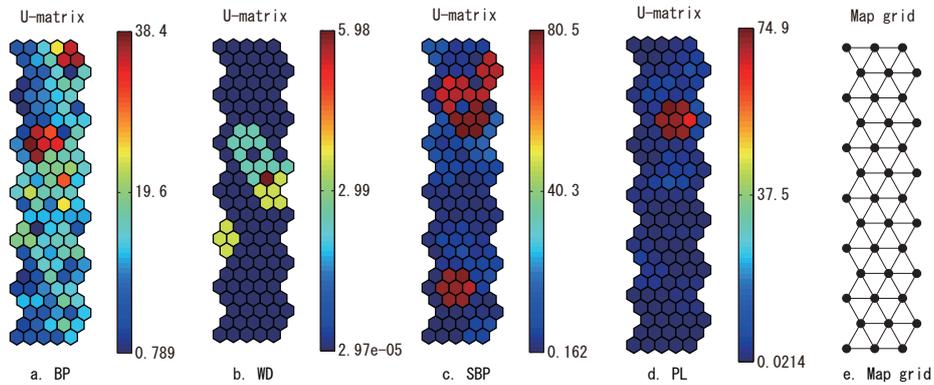


図 27: 学習完了後の重み (入力層-中間層間) を U-Matrix で示したもの

重み減衰を用いた誤差逆伝播法

WD では 1 番目の入力のみが値を持っていると言える結果を示しており、よって 1 番目の入力が必要であるという解釈ができた。

潜在性を用いない潜在学習法

SBP では各重みでそれぞれの重みをもっており変数の重要性について傾向などを確認することは困難であった。

潜在学習法

PL では 1 番目の入力のみが潜在性を持っており、よって 1 番目の入力が必要であるという解釈ができた。

ロジスティック回帰分析 (変数削減あり)

Logi Reg の結果であるが、変数選択を実施して解析を行ったところ、1 番目の変数が選択された。よって 1 番目の入力が必要であるという解釈ができた。

ロジスティック回帰分析 (変数削減なし)

Logi Reg ALL の結果であるが、変数 1 の方が変数 2 よりも係数が大きく、よって 1 番目の入力が必要であるという解釈ができた。

という結果を得た。以上の結果をまとめると、当解析で用いた六つの解析手法のうち変数についての解釈が可能であったのは、重み減衰を用いた誤差逆伝播法、潜在学習法、変数削減を用いたロジスティック回帰分析、ロジスティック回帰分析の四つであり、これらはどれも重要変数について正解できた。正解がわかっているデータを用いたこの解析結果より、潜在学習法の変数抽出結果の妥当性が確認できた。

3.2.3 学習第 1 段階が学習第 2 段階に与える効果の検証

ここで、潜在学習法の学習第 1 段階で自己組織化マップを使用することについて、学習第 2 段階の学習後の重みを U-Matrix 形式で表示させ確認を行った (図 27 に結果を示す)。この図は誤差逆伝播法、重み減衰を用いた誤差逆伝播法、潜在性を用いない潜在学習法 (自己組織化マップを事前学習に用いた誤差逆伝播法)、潜在学習法のそれぞれの手法の学習完了後の重みの状態である。図 27 (e) には U-Matrix に対応するニューロン構造を示している。本来はクラスタの状態を確認するための用途で使用される U-Matrix であるが、ここでは重みの状態を確認する目的で使用している。図 27 (a) で示す誤差逆伝播法は重みの初期値をランダムに割り当てたものであり、学習後の重みもランダムであることを示している。(b) に示す重み減衰を用いた誤差逆伝播法は重みの初期値はランダムに割り当てられているが、学習過程で重みの減衰が起こることで、特定のニューロンの重みだけ大きく、それが学習に使用されていることが確認できる。(c) の自己組織化マップを事前学習に用いた誤差逆伝播法は、重みの初期

値にランダムではなく自己組織化マップより得られた重みを使用している。このためランダムであれば (a) のような結果になる重みに特徴が見られる。この特徴は特定のニューロンのみ大きな重みが割り当てられていることより生じており、このことが図 21 に示した精度につながっていると考えられる。最後に (d) に示す潜在学習法の重みであるが、(c) の結果よりもさらに特定のニューロンのみ大きな重みが存在していることが確認できる (この際の重み (U-Matrix にて示さない場合) については図 51 に結果を示す)。重みに偏りが確認できた (c) と (d) であるが、これらは同じく偏りが確認できた (b) とは傾向が異なる ((b) ではマップ中央付近、(c) と (d) ではマップ上部と下部)、そのため (c) と (d) の結果は自己組織化マップの結果を初期値とすることで得られたもので、自己組織化マップの効果であると判断できる。この図 27 の結果より、自己組織化マップにより作成された重みは潜在学習法の学習第 2 段階に特徴が残っていることが確認でき、これにより自己組織化マップ (潜在学習法学習第 1 段階) を誤差逆伝播法 (潜在学習法学習第 2 段階) の事前学習として用いることの効果が確認できたと結論付ける。

3.2.4 変数着目効果の検証

ここでは潜在学習法の特徴である「分散の大きな変数への着目」の着目度合の影響を検証する。表 2 は各パラメータ r 時点の各変数の潜在性を示したものである。これによると変数 1 (X1) はパラメータ値が 1.0 から 2.0 の時まで一貫して最高値の 1.0000 となっていることが分かる。しかし変数 2 (X2) はパラメータ値が 1.0 の際は 0.0410 であり、これから徐々に数値は下がりはじめ、1.7 の時点で 0.000 となっている。式 (2-48) より、潜在性の値により求められる距離に重みがつけられるため、表 2 の結果はパラメータ r が増加するにしたがい、着目する変数と着目されない変数で解析に使用される度合いが変化していることを示すものである。この場合、変数 2 はパラメータ r の増加により、より活用されなくなっていることを意味し、パラメータ値が 1.7 から 2.0 の段階では変数 1 のみが解

析に使用されている状態となっている。

この潜在性の変化とモデル汎化能力の変化を示したものが図 28 である。誤差逆伝播法よりも、自己組織化マップを事前学習に用いた誤差逆伝播法の方が汎化能力は高く、そして潜在学習法はより良い汎化能力を示している。パラメータ r により汎化能力の変化は見られるものの、どのパラメータ値でも潜在学習法の結果が良いものであった。これは解析に活用する変数が絞られることにより得られる結果であり、「着目した変数を活用する」という潜在学習法は期待通りの動作をしていることが確認できた。

表 2: パラメータ r の変化と各変数の潜在性

パラメータ r	X1	X2
1.0	1.0000	0.0410
1.1	1.0000	0.0266
1.2	1.0000	0.0154
1.3	1.0000	0.0090
1.4	1.0000	0.0047
1.5	1.0000	0.0014
1.6	1.0000	0.0001
1.7	1.0000	0.0000
1.8	1.0000	0.0000
1.9	1.0000	0.0000
2.0	1.0000	0.0000

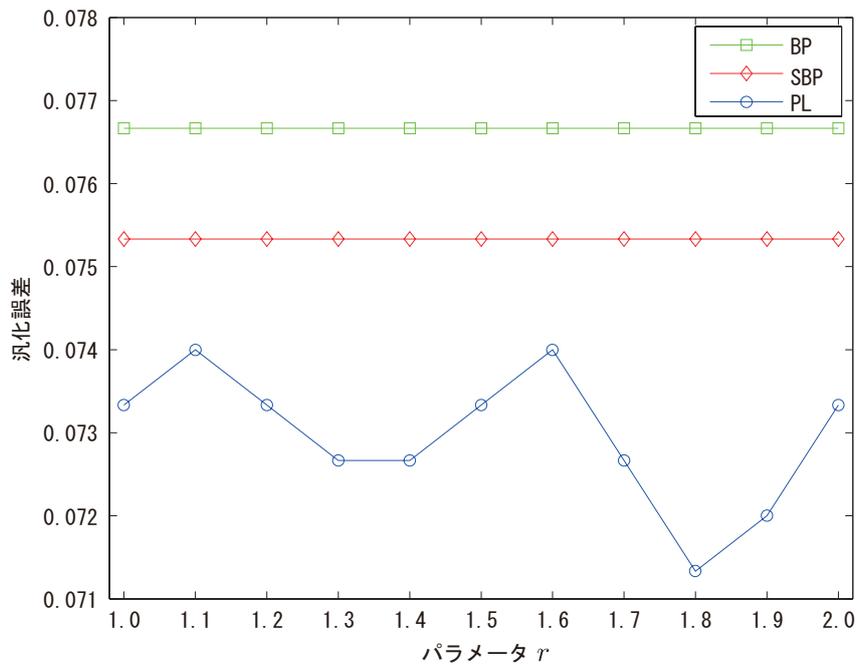


図 28: パラメータ r の変化と汎化能力

3.2.5 初期重みをランダムに設定した場合の検証

当解析では自己組織化マップの初期重みを「線形な初期化」により設定したが、初期値をランダムに設定した場合でも潜在学習法の有効性を確認した。参考までに学習前の重みとデータとの関係を図 29 に、学習後の重みとデータとの関係を図 30 に、モデル汎化能力を図 31 (10 個のモデル中で最も良い汎化能力を示した際の結果 (潜在性を用いない潜在学習法: 0.0533、潜在学習法: 0.0467 (パラメータ $r = 1.7$)) に、SBP の BP 部で使用された自己組織化マップの重みを図 32 に、潜在学習法の学習第 2 段階で使用された学習第 1 段階の重みを図 33 に、SBP および潜在学習法の学習完了時の重みを図 34 に、そして図 34 を U-Matrix にて表示したものを図 35 に示す。

自己組織化マップの重み初期値をランダムに設定した場合でも予測精度は SBP より潜在学習法の方が高く (図 31)、重要変数も潜在学習法により変数 1 が重要と判断され (図 34)、また、学習結果を確認すると SBP より PL の方が明確であるという結果を得たことから (図 32 と図 33)、自己組織化マップの重み初期値をランダムに設定した場合でも潜在性が活

かされていることが確認できた。

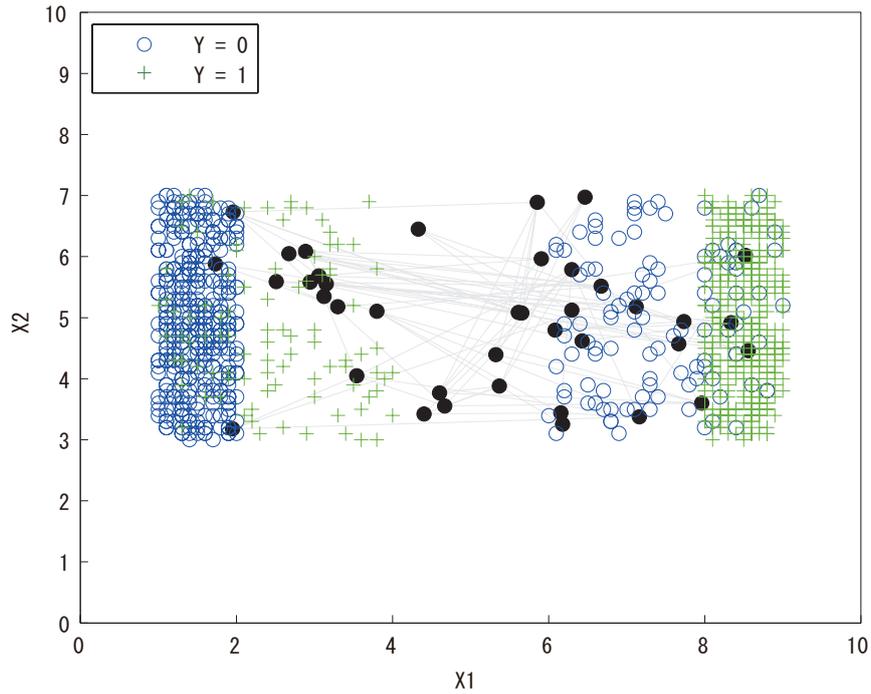


図 29: 人工データの散布図と自己組織化マップの出力ニューロン配置との関係 (学習前) (初期重みランダム設定)

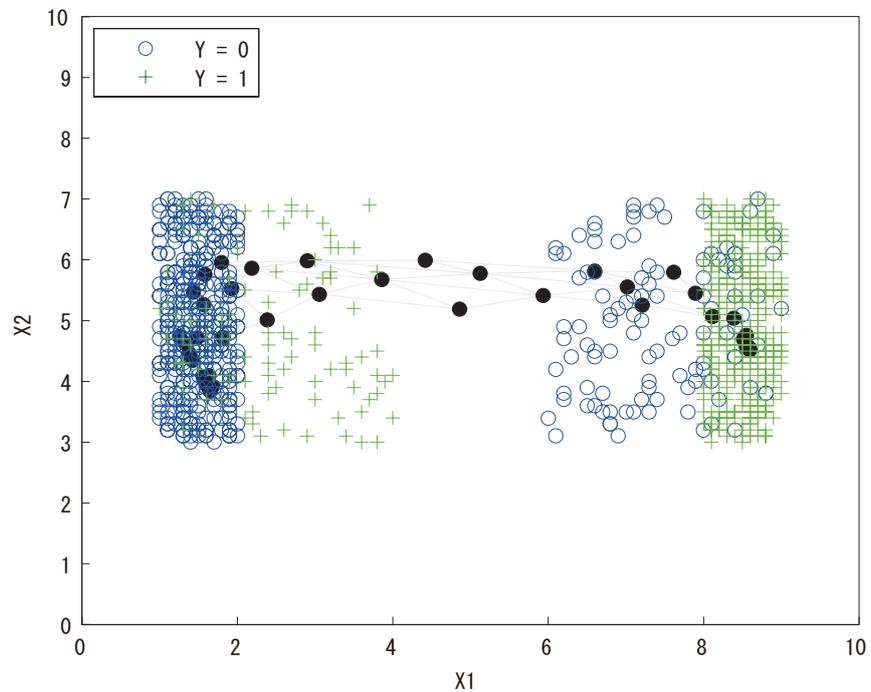


図 30: 人工データの散布図と自己組織化マップの出力ニューロン配置との関係 (学習後) (初期重みランダム設定)

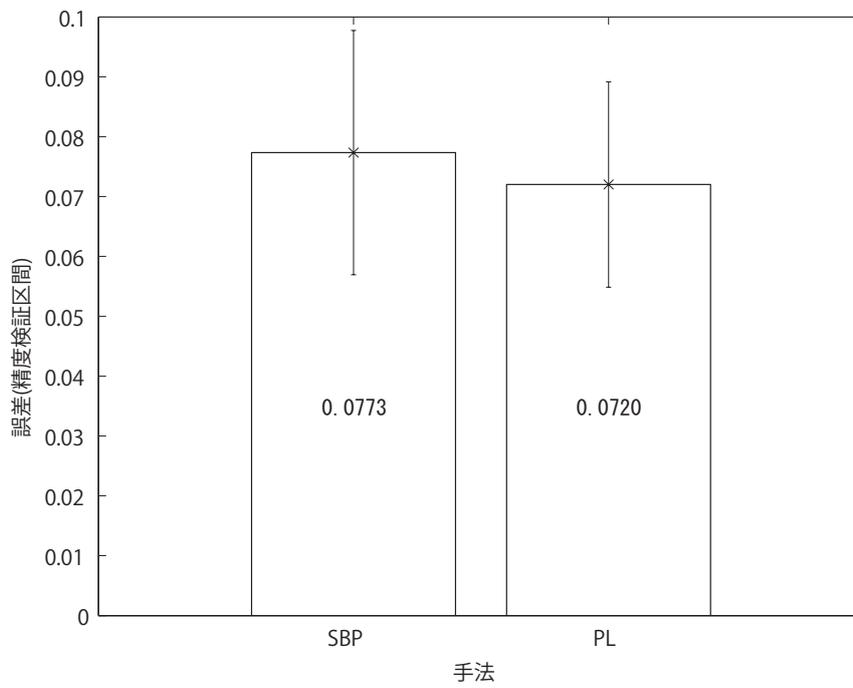


図 31: モデル汎化能力 (初期重みランダム設定)

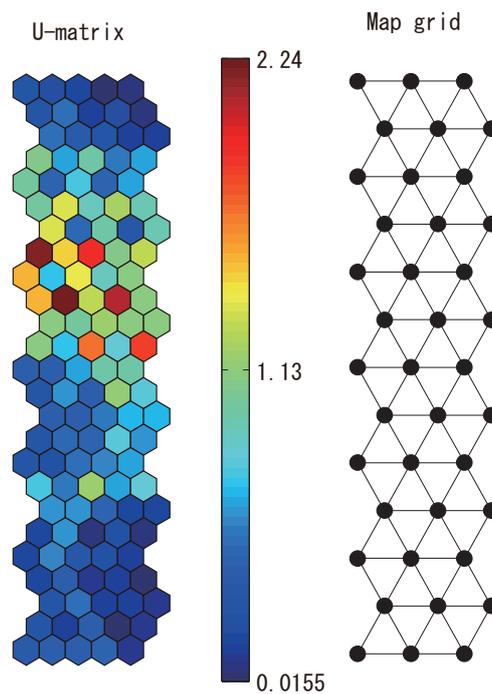


図 32: SBP における自己組織化マップの学習結果 (初期重みランダム設定)

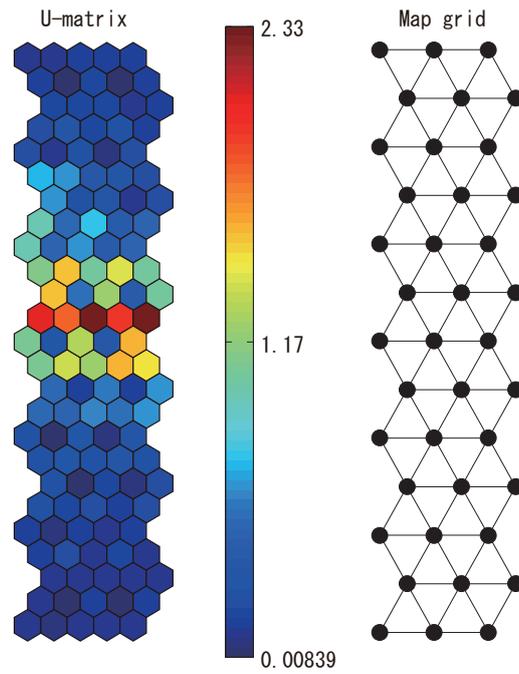


図 33: PL における自己組織化マップの学習結果 (初期重みランダム設定)

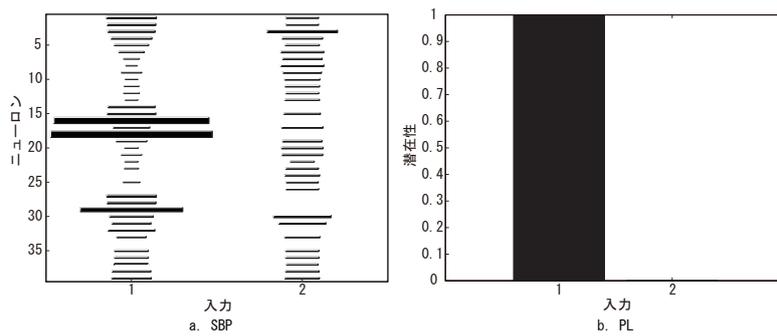


図 34: 重み (入力層-中間層) (初期重みランダム設定)

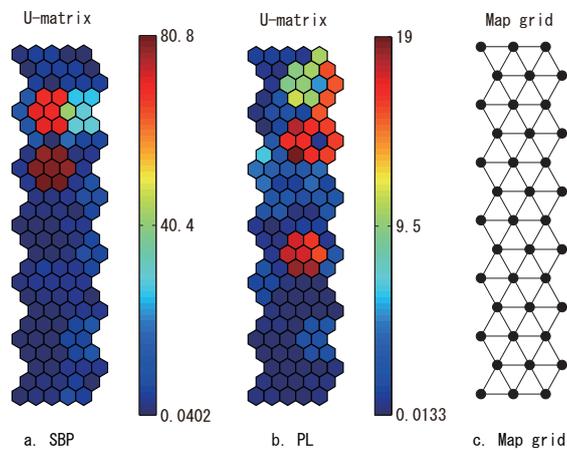


図 35: 学習完了後の重み (入力層-中間層間) を U-Matrix で示したもの (初期重みランダム設定)

表 3: 人工データ (高頻度ノイズデータ) (一部表示)

X1	X2	Y	Flg
1.9	6.0	0	0
1.4	6.6	0	0
1.2	3.5	0	0
1.5	6.9	0	0
6.4	5.8	0	1
～略～			
8.1	5.4	1	0
2.8	3.3	1	1
2.7	3.4	1	1
8.2	4.4	1	0
2.2	6.0	1	1

3.3 ノイズ混入率が高いデータを用いた検証

ここでは表 1 で用いたデータのノイズ混入率をあげ、正解となるデータサンプルが少ない状態のデータを用いて各手法の解析結果を検証する。これは表 1 のデータを複雑にさせたデータと考えることができる (データの具体例は表 3 に示す。本データではノイズ発生のトリガとなる Flg「1」は約 50%の確率で発生する。また図 36 に散布図を示す)。

はじめに人工データの解析結果より潜在学習法の汎化能力について確認を行う。使用したネットワークの構成 (BP・WD・SBP の BP 部・PL の学習第 2 段階) は図 37 に示すものである。このネットワークは中間ニューロンが 40 個、出力ニューロンが 2 個から構成されたネットワークである。

表 3 に示すデータを誤差伝播法 (BP)、重み減衰を用いた誤差逆伝播法 (WD)、誤差逆伝播法の事前学習に自己組織化マップを用いた手法 (潜在性を用いない潜在学習法) (SBP)、潜在学習法 (PL)、変数削減を使用したロジスティック回帰分析 (Logi Reg)、ロジスティック回帰分析 (Logi Reg ALL) にてそれぞれ解析した結果を図 38 に示す。なお、解析データは 1000 件の 70%にあたる 700 サンプルをモデル学習用、15%にあたる 150 サンプルを過学習を抑制するための早期停止用、15%にあたる 150 サンプルをモデル汎化能力測定用として用いた。さらにこの分割比率は固定

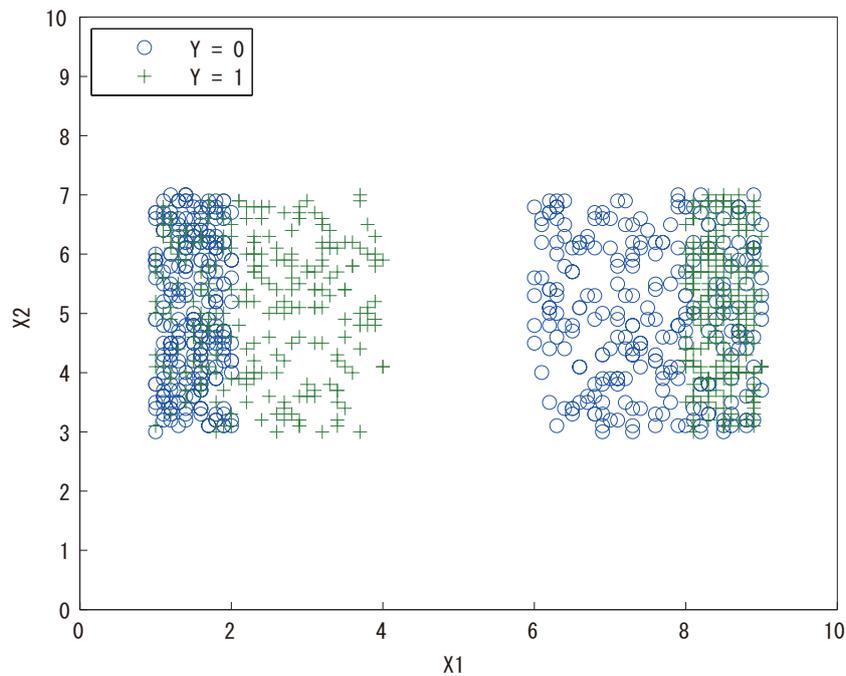


図 36: 人工データの散布図 (高頻度ノイズデータ)

とし、各用途にて使用されるデータサンプルをランダムに変更して 10 パターンのデータを作成した。これは良い結果をもたらすデータサンプルが偶然に使用されることを防止するためである。よって結果は 10 個のモデルから得られた平均値を示している。図 38 に示す結果より、人工データ解析モデルの汎化能力は高いものから、PL・WD・SBP・BP・Logi Reg ALL・Logi Reg の順であることが確認できる。

ここで自己組織化マップの結果を確認する。図 39 は潜在性を用いない潜在学習法での分類結果である。データはマップの中心を境として上下二つのクラスが形成されていることが確認できる。一方、図 40 は潜在学習法での分類結果である、こちらもマップの上下でクラスが形成されていることが確認できる。しかし、暖色の境界線が複雑になっており、分割がより強力に行われていることが確認できる。

続いて重要変数の抽出について確認を行う。図 41 は各変数の重要性解釈のために必要な情報を示したものである。図 41 は入力層-中間層間の重み (図 41 (a) から (c))、潜在性 (図 41 (d))、ロジスティック回帰分析の係数 (図 41 (e) と (f)) を示している。これら図 41 の各図は 10 個のモデル中で最も良い汎化能力を示した際の結果を示している (誤差逆伝播法:

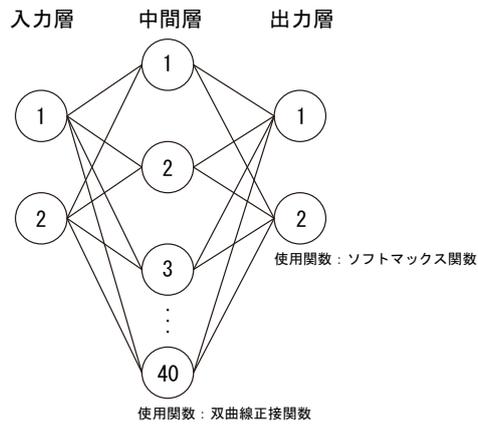


図 37: ネットワーク構成 (高頻度ノイズデータ)

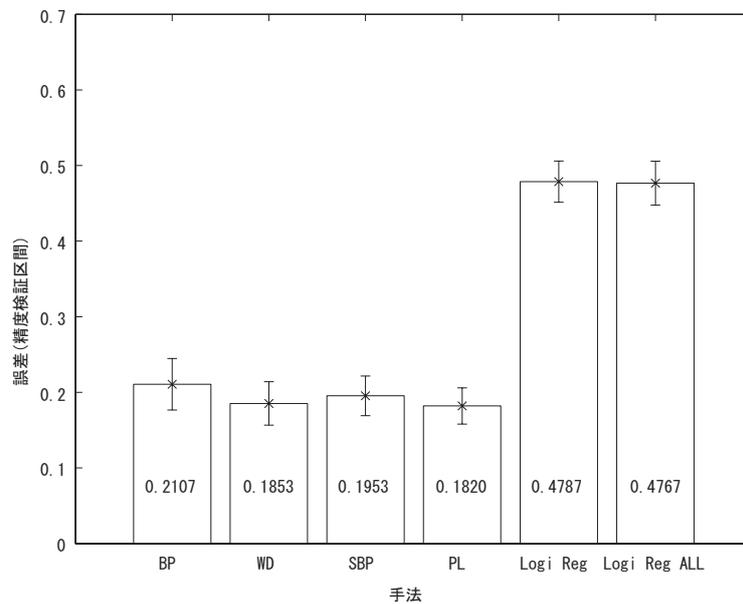


図 38: モデル汎化能力 (高頻度ノイズデータ)

0.1667、重み減衰を用いた誤差逆伝播法 : 0.1533 (パラメータ $\lambda = 0.9$)、潜在性を用いない潜在学習法 : 0.1533、潜在学習法 : 0.1400 (パラメータ $r = 1.4$)、ロジスティック回帰分析 (変数選択あり) : 0.4733、ロジスティック回帰分析 (変数選択なし) : 0.4733)。図 41 の各結果より

誤差逆伝播法

BP の重みは各重みでそれぞれの重みをもっており変数の重要性について傾向などを確認することは困難であった。

重み減衰を用いた誤差逆伝播法

WD では 1 番目の入力の方が 2 番目と比較して大きな値を持ってお

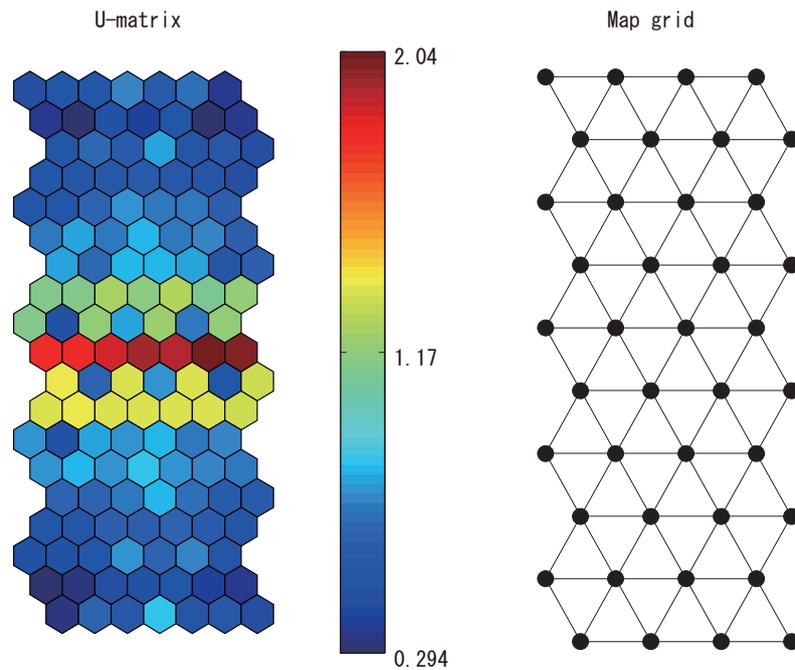


図 39: SBP における自己組織化マップの学習結果 (高頻度ノイズデータ)

り 1 番目の入力が必要であるという傾向を確認することができた。

潜在性を用いない潜在学習法

SBP では各重みでそれぞれの重みをもっており変数の重要性について傾向などを確認することは困難であった。

潜在学習法

PL では 1 番目の入力が 2 番目の入力よりも大きな潜在性を持っており 1 番目の入力が重要であるという解釈ができた。

ロジスティック回帰分析 (変数削減あり)

Logi Reg の結果であるが、変数選択を実施して解析を行ったが、変数削減は行われなかった結果であった。重要度については 2 番目変数 (X2) の係数が 1 番目変数 (X1) の係数よりも大きな値を持っており、これより 2 番目変数が重要であると解釈ができた。

ロジスティック回帰分析 (変数削減なし)

Logi Reg ALL の結果であるが、係数は 2 番目変数 (X2) の方が 1 番目変数 (X1) よりも大きな値を持っており、これより 2 番目変数が重要であると解釈ができた。

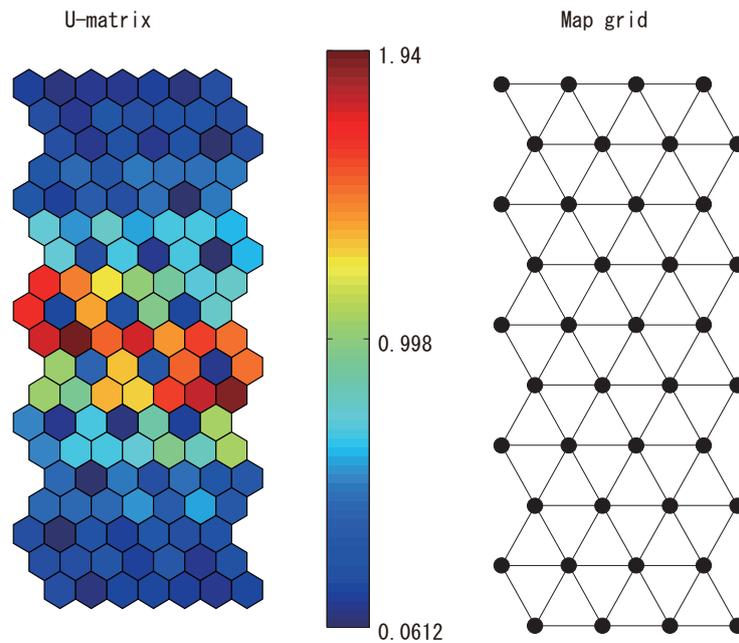


図 40: PLにおける自己組織化マップの学習結果 (高頻度ノイズデータ) という結果を得た。以上の結果をまとめると、当解析で用いた六つの解析手法のうち変数についての解釈が可能であったのは、重み減衰を用いた誤差逆伝播法、潜在学習法、変数削減を用いたロジスティック回帰分析、ロジスティック回帰分析の四つであり、そして重要変数について正解できた手法は重み減衰を用いた誤差逆伝播法と潜在学習法であった。この結果より、潜在学習法の変数抽出結果の妥当性が確認できたと考えることができる。また、ロジスティック回帰分析は汎化能力 (図 38) と重要変数抽出の結果 (図 41 (e) と (f)) より、複雑なデータの解析に対応できていないと言える結果であった。

またここで、U-Matrixにて学習後の重みを確認すると (図 42 に結果を示す)、図 42 (a) で示す誤差逆伝播法は学習後の重みがランダムであることを示している。(b) に示す重み減衰を用いた誤差逆伝播法では、数個のニューロンの重みだけが大きい値を持っていることが確認できる。(c) の誤差逆伝播法の事前学習に自己組織化マップを用いた手法では (b) よりもさらに少ない個数のニューロンのみが大きな重みを持っていることが確認できる。最後に (d) に示す潜在学習法の重みであるが、こちらも特定のニューロンのみに重みが存在していることが確認できた。よって、

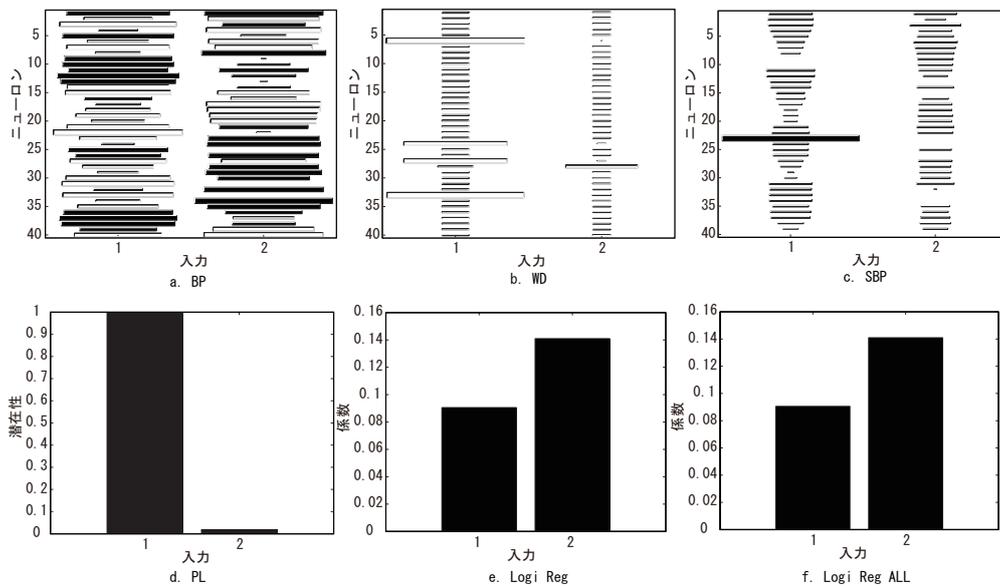


図 41: 重み (入力層-中間層) (高頻度ノイズデータ)

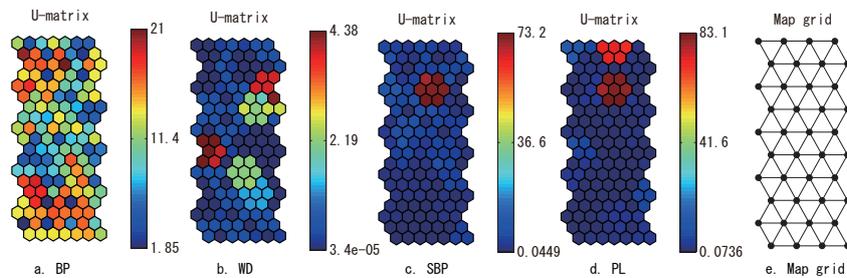


図 42: 学習完了後の重み (入力層-中間層間) を U-Matrix で示したもの (高頻度ノイズデータ)

自己組織化マップ (潜在学習法学習第 1 段階) を誤差逆伝播法 (潜在学習法学習第 2 段階) の事前学習として用いることの効果が確認できた。

続いて各パラメータ r での各変数の潜在性を確認する (表 4 に結果を示す)。この表によると変数 1 (X1) の潜在性が各パラメータ r の値を通して 1.0000 であり、変数 2 (X2) はパラメータ r が増加するにしたがい減少していることが確認できる。パラメータが 1.8 以降は変数 2 の潜在性は 0.0000 となっており、すなわち、変数 1 のみが活用される状態になっている。潜在学習法は分散という軸から活用する変数を決定するため、ノイズが高頻度であってもその影響を避けて重要変数を活用していることが確認できた。

最後にパラメータ r と汎化能力の関係を確認する。図 43 に示すグラフ

より、誤差逆伝播法、誤差逆伝播法の事前学習に自己組織化マップを用いた手法、潜在学習法の順で汎化能力が向上していることが確認できる。これにより、潜在性によって使用する変数の度合いを調整するという潜在学習法の特徴が汎化能力に活かしていることが確認できた。

表 4: パラメータ r の変化と各変数の潜在性 (高頻度ノイズデータ)

パラメータ r	X1	X2
1.0	1.0000	0.0818
1.1	1.0000	0.0625
1.2	1.0000	0.0447
1.3	1.0000	0.0294
1.4	1.0000	0.0186
1.5	1.0000	0.0088
1.6	1.0000	0.0024
1.7	1.0000	0.0003
1.8	1.0000	0.0000
1.9	1.0000	0.0000
2.0	1.0000	0.0000

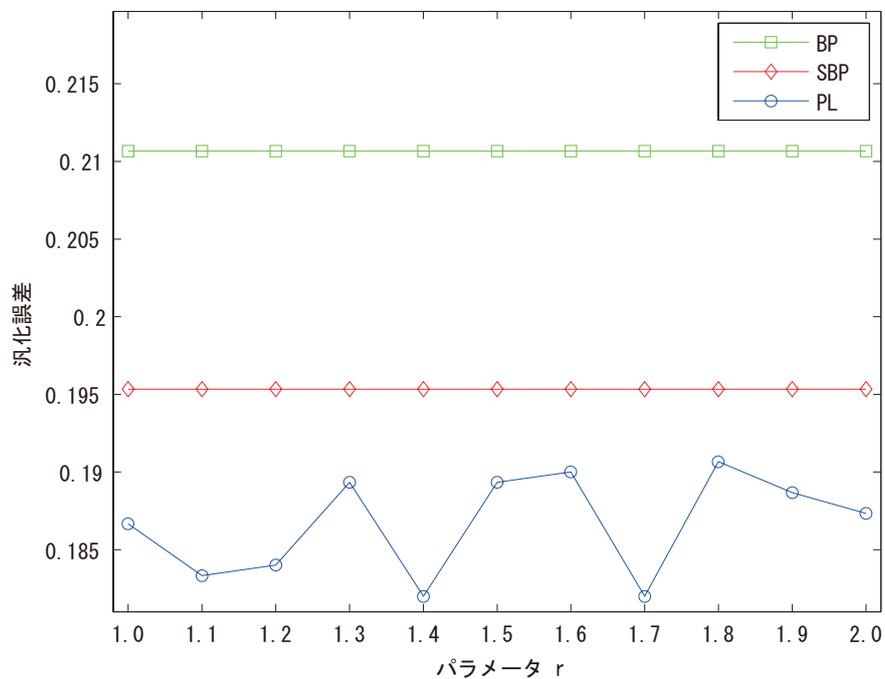


図 43: パラメータ r の変化と汎化能力 (高頻度ノイズデータ)

3.4 多重共線性発生データに対する解析

多変量解析は多重共線性問題が生じているデータの解析が困難であると 1.5.2 で述べた。ここでは多重共線性が生じているデータを誤差伝播法 (BP)、重み減衰を用いた誤差逆伝播法 (WD)、誤差逆伝播法の事前学習に自己組織化マップを用いた手法 (潜在性を用いない潜在学習法) (SBP)、潜在学習法 (PL)、変数削減を使用するロジスティック回帰分析 (Logi Reg)、ロジスティック回帰分析 (Logi Reg ALL) を用いて解析し解析結果について手法による違いを確認していく。使用したネットワークの構成は中間ニューロンが 39 個、出力ニューロンが 2 個から構成されたネットワークである。

当検証で用いたデータは表 5 に示すものである。このデータについて散布図を図 44 に示す。このデータは 3 変数、1000 サンプルから構成されているデータで、変数 1 (X1) と変数 2 (X2) を表 1 と同様の方法で作成し、これに変数 1 の 1/2 の値を変数 3 (X3) として加えたものである。この変数 1 と変数 3 の相関係数は 1 であり多重共線性が発生しているデータである。また、表 5 中の Flg であるがこれが「1」になる確率は約 50% である。なお、正解となる重要変数は表 1 と変わらず変数 1 である (変数 3 からも Y の予測が可能のため変数 3 も正解ではあるが、値が変数 1 よりも小さいため正解データを変数 1 とした)。

表 5: 多重共線性が生じている人工データ (多重共線性データ) (一部表示)

X1	X2	X3	Y	Flg
1.5	6.8	0.75	0	0
8.8	5.8	4.4	0	1
8.6	3.9	4.3	0	1
1	6	0.5	0	0
7.1	6.7	3.55	0	1
～略～				
8.3	5.6	4.15	1	0
2.2	6.7	1.1	1	1
8.5	6.6	4.25	1	0
1.3	4.1	0.65	1	1
8.5	6.3	4.25	1	0

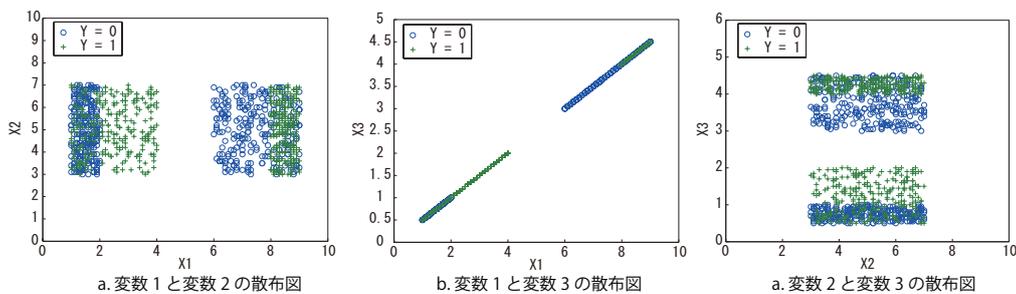


図 44: 人工データの散布図 (多重共線性データ)

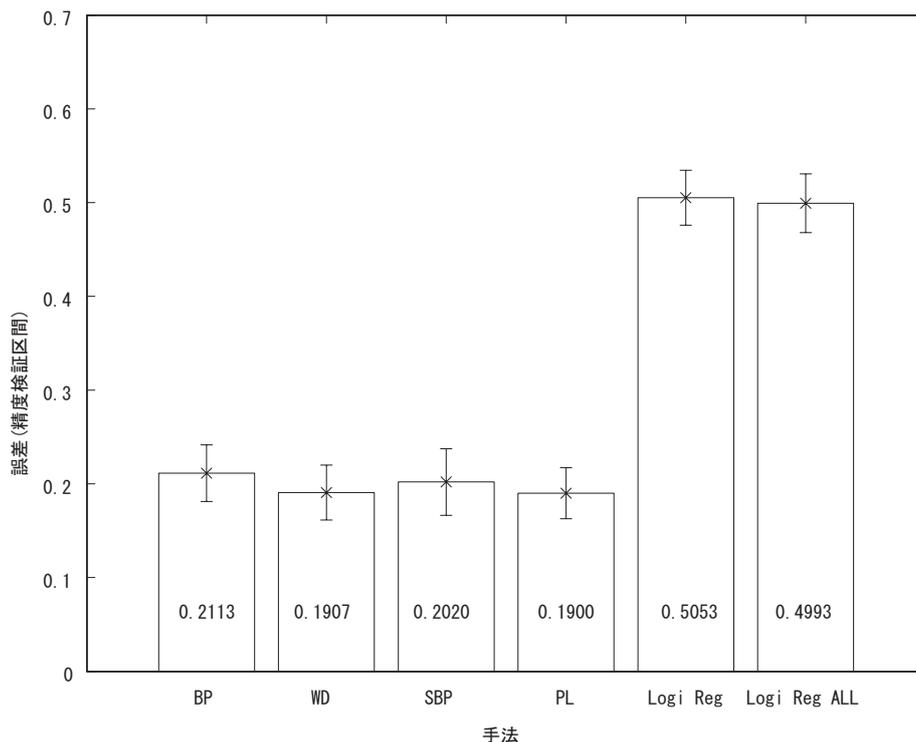


図 45: モデル汎化能力 (多重共線性データ)

図 45 はモデルの汎化能力 (10 個のモデルの平均値) を示したものである。モデルは汎化能力が高い順に、PL・WD・SBP・BP・Logi Reg ALL・Logi Reg という結果であった。ここで自己組織化マップの結果を確認する。図 46 は潜在性を用いない潜在学習法での分類結果である。データはマップの中心より上下に分割されていることが確認できる。一方、図 47 は潜在学習法での分類結果である、こちらもマップの上下でクラスが形成されていることが確認できる。しかし、図 46 では色がまだらになっていたものが均一化しており、分割がより強力に行われていることが確認できる。

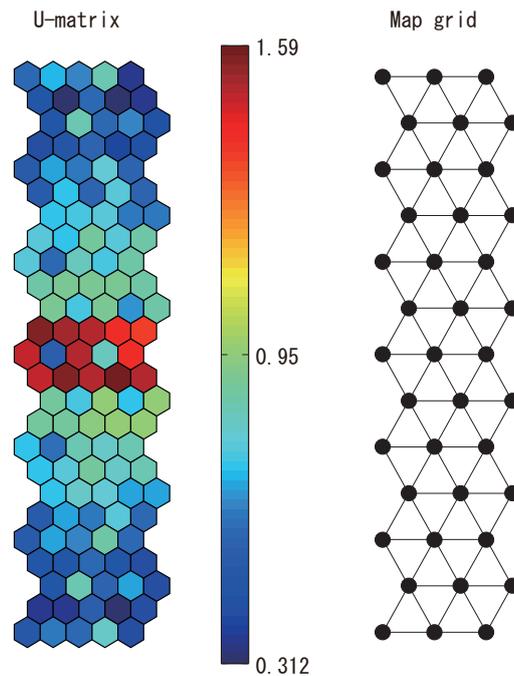


図 46: SBP における自己組織化マップの学習結果 (多重共線性データ)

続いて重みについて解釈を行う。図 48 の各図は 10 個のモデル中で最も良い汎化能力を示した際の重みを示している (誤差逆伝播法 : 0.1733、重み減衰を用いた誤差逆伝播法 : 0.1467 ($\lambda = 0.9$)、潜在性を用いない潜在学習法 : 0.1600、潜在学習法 : 0.1533 (パラメータ $r = 1.9$)、ロジスティック回帰分析 (変数選択あり) : 0.5400、ロジスティック回帰分析 (変数選択なし) : 0.4933)。図 48 の各図より

誤差逆伝播法

各ニューロンがそれぞれの値を持っており重みの傾向を確認することが困難な結果であった。

重み減衰を用いた誤差逆伝播法

各ニューロンがそれぞれの値を持っているが、変数 1 と変数 3 の重みの変数 2 よりも大きいため、変数 1 と変数 3 がネットワークで活用されていると解釈ができた。ただし、変数 1 と変数 3 のどちらがより重要であるかという解釈は困難であった。

潜在性を用いない潜在学習法

各ニューロンがそれぞれの値を持っており重みの傾向を確認するこ

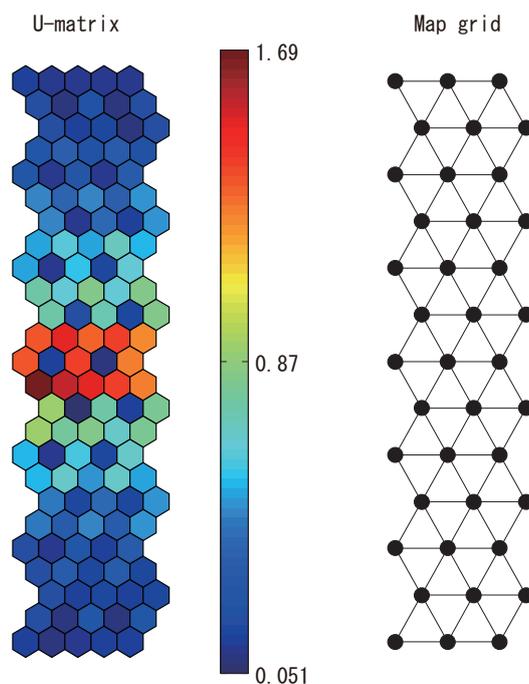


図 47: PL における自己組織化マップの学習結果 (多重共線性データ)

とが困難な結果であった。

潜在学習法

潜在性を確認すると、変数 1 が最も潜在性を持っており、これが重要変数であると解釈ができた。

ロジスティック回帰分析 (変数削減あり)

変数削減により変数 3 のみが選択された。このため重要変数は 3 番目の変数であると解釈ができた。

ロジスティック回帰分析 (変数削減なし)

係数を確認すると変数 2 の係数が最も大きく、これより変数 2 が重要変数であると解釈ができた。しかしここで変数 3 の係数を確認すると、これは 0 であった。変数 1 の値を $1/2$ したものが変数 3 の値であり、変数 1 と同様にターゲット予測に役立つ変数であるため、この係数は不可解なものである。これはデータの多重共線性が問題で係数の値がおかしくなっているためであると考えられる。

という結果であった。

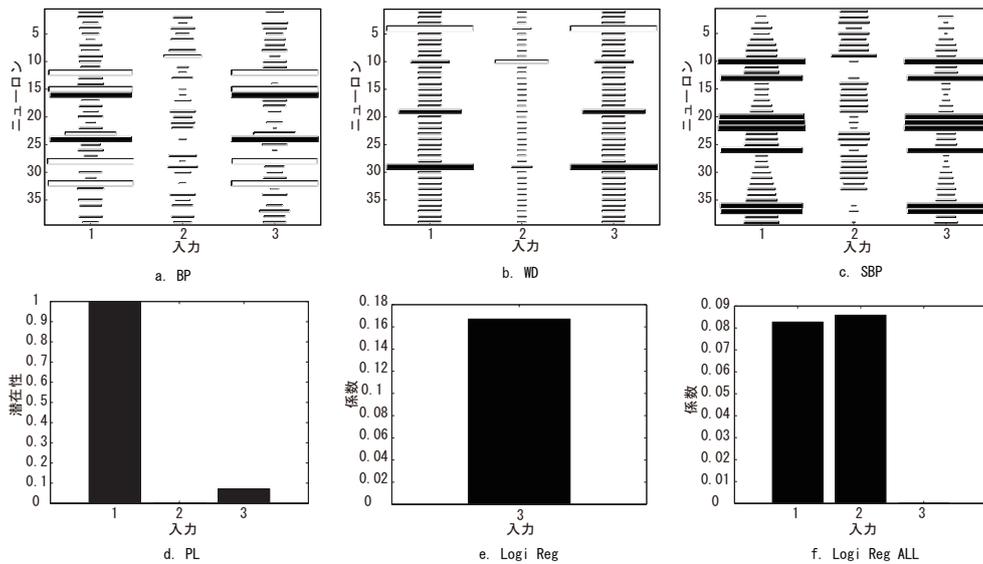


図 48: 重み (入力層-中間層) (多重共線性データ)

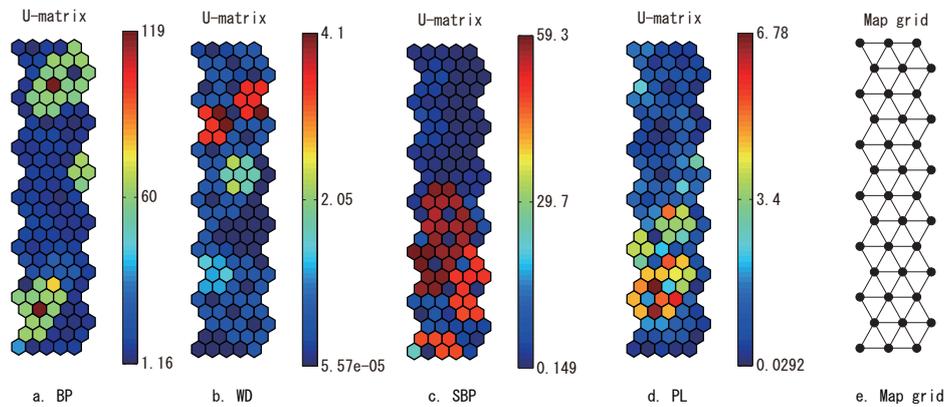


図 49: 学習完了後の重み (入力層-中間層間) を U-Matrix で示したものの (多重共線性データ)

当解析では多重共線性が発生しているデータの解析を行った。変数について解釈できた手法は重み減衰を用いた誤差逆伝播法、潜在学習法、ロジスティック回帰分析 (変数選択あり) とロジスティック回帰分析 (変数選択なし)、の4手法であった。しかしロジスティック回帰分析 (変数選択なし) は多重共線性の影響を受けて係数が不可解であり信頼性は低い結果であった。一方、ロジスティック回帰分析 (変数選択あり) は変数選択により3番目の変数のみが選択され解析に使用された。変数3は変数1の1/2の値でありこれを使用してもターゲットの予測を行うことができる。データ設計上の重要変数は変数1であるものの、有効な変数を選択して使用するという点では妥当性のある結果であった。これと比較して潜在

学習法は変数1の潜在性が最も大きく、ついで変数3の潜在性が大きな値を持っているという結果であり、最も妥当性のある結果となった。重み減衰を用いた誤差逆伝播法は重要変数の解釈は可能であったが、変数1と変数3のどちらがより大切かというところまでは判断ができない結果であった。当解析により、潜在学習法は多重共線性が発生しているデータであっても解析が可能であり、変数解釈も妥当性のある結果を得ることができ、また、重要な度合いまでも判断できる手法であると確認できた。

続いてU-Matrixにて学習後の重みを確認する(図49に結果を示す)。図49(a)で示す誤差逆伝播法は特定の重みが大きな値を持っていることが確認できる。(b)に示す重み減衰を用いた誤差逆伝播法では、数個のニューロンの重みだけが大きい値を持っていることが確認できた。(c)の誤差逆伝播法の事前学習に自己組織化マップを用いた手法ではU-Matrixの下部に暖色のニューロンが集まっており、特定のニューロンが大きな重みを持っていることが確認できる。最後に(d)に示す潜在学習法の重みであるが、こちらもU-Matrixの下部に大きな値を持つニューロンが集まっていることが確認できる。当解析ではこれまでの結果とは異なり誤差逆伝播法の重みもランダムにはなっておらず特徴が確認できるものであった。しかし重みの偏り具合は異なっており、誤差逆伝播法がU-Matrix上下に大きな重みがあるのに対して、誤差逆伝播法の前学習に自己組織化マップを用いた手法と潜在学習法ではU-Matrix下部に大きな重みが集中していた。図49に示している結果はターゲット(Y)を予測するために最適化された結果であり、またU-Matrixも本来の使用方法ではなくここでは偏りの傾向しか比較できないが、この偏りの違いが図45に示す予測精度に影響を与えているため自己組織化マップを予測の際の初期重み作成に使用することの有効性が確認できた。

最後にパラメータ値と潜在性の関係であるが、表6に結果を示している。これによるとパラメータ値が1.6の段階で変数2(X2)は全く解析に活用されなくなり、そして変数3(X3)はパラメータ値が2.0の際においても少しではあるが活用されている状態にあることが確認できる。この動作は実データの解析において「2番目に重要な変数を目的変数予測のため

めに少し活用する」という動作を示したものと言え、着目度 (潜在性) によって変数の活用を変化させるという潜在学習法ならではの動作が確認できた。各パラメータ値と汎化能力の関係は図 50 に示している。誤差逆伝播法よりも、誤差逆伝播法の事前学習に自己組織化マップを用いた手法、そして潜在学習法の順で汎化能力が高いことが確認できる。

ところで、当解析では変数削減を使用したロジスティック回帰分析は変数 3 を選択した。変数 1 を選択しても変数 3 を選択しても予測は可能なため、どちらを選択しても正解であると言えるが、このことは本データのように素性がわかっている人工データであるため言えることである。実際のデータではどの変数が重要かわからないため、当解析で実施したように統計上の基準により機械的に変数を取り除くことは好ましくないと見える。一方、潜在学習法も変数削減を行っているが、これは削除ではないため (重みが非常に小さいなどの理由で結果的に使用されていないだけである)、解析結果より変数の重要性・重要度合いをマーケティングが確認 (すなわち、マーケティング仮説上の基準から確認) しながらその変数を解析に使用すべきかどうか判断できる。この、変数を解析から実際に取り除かず、そして重要度合い情報を出力する機能は潜在学習法の強みでありマーケティング分析に有効であると考えられる。

3.5 構造学習との違い

ここでは潜在学習法の変数削減機能について、変数削減を実施する他のニューラルネットワーク手法と比較を行う。潜在学習法は潜在性に着目して学習を行うが、このとき潜在性が低い変数は大きな重みを持たないため (学習第 2 段階の初期重みに学習第 1 段階で得られた重みと潜在性を乗算した値を設定するため) 変数の削減を行って解析していることと同義となる (図 51 に図 22 (d) 時点の重みを示す)。しかしこの削減は図 22 (d) および (e) に示すようにロジスティック回帰分析が完全に変数を削減しているのに対して潜在学習法では変数そのものは削減していない。このことより潜在学習法は柔らかい変数削減を実施する手法であるといえる。

変数削減を実施するニューラルネットワーク手法は数多く存在するが [26], [27], [28], [29], [30], [31], [32]、ここでは構造学習と潜在学習法を比較する。構造学習とは忘却付き構造学習を提案した石川によれば「ネットワークの構造や隠れユニット¹の意味づけを重視する学習」 [31] を意味し、これはすなわち本論文で目標とする「解釈可能なニューラルネットワーク」と目的を同じとするものである。構造学習は数多く提案されており、本章の解析でも使用した重み減衰や Le Cun らの Optimal Brain Damage (OBD) [26] などがあげられる。しかしこれら、例えば OBD はネットワークの構造を探る目的で構造学習を実施するのではなく過学習の抑制、学習速度向上、汎化能力向上などを期待している [26] ものである。これに対して忘却付き構造学習は目的を「データ中に含まれる規則性を反映した単純な構造のネットワークを求めること」 [31] としている。忘却付き構造学習では誤差評価関数に $E(\mathbf{w}) + \varepsilon' \|\mathbf{w}\|$ を使用し学習を実施する。式中の $\|\mathbf{w}\|$ は重みの絶対値和であり (L1 正則化と呼ばれている)、これを使用することで多くの重みを 0 とすることができる。また ε は忘却に関する係数である。

構造学習と潜在学習法は「データ構造を得る」という目的は同じとしているものの、解析処理の平易さで大きく異なっていることが強調できる。構造学習は前述したようにロジックが誤差評価に組み込まれている (忘却付き構造学習は L1 正則化、重み減衰は L2 正則化) のに対し、潜在学習法では潜在性を算出して学習第 1 段階で重み付き距離を基に学習を行うことと、潜在性を学習第 2 段階の初期重み決定処理に使用するのみであり、ロジックが平易である。この手法の平易さは計算処理の容易性、既存手法への高い拡張性という効果を持つとともに、第三者への解析手法解説でも活かされる。マーケティング分析では解析結果だけが重要なのではなく、データ解析依頼元といった第三者に手法を説明することも重要であることを 1.6.2 にて述べたが、潜在学習法ではこれが容易であり、これらの点で構造学習と比べて優位であると言える。

また、構造学習と潜在学習法はデータ構造獲得の際の動作の考え方も

¹本論文ではこれを中間ニューロンと表現している。

異なっている。構造学習ではL1またはL2の正則化を用いて大きい重みを小さくしようと動作する。一方、潜在学習法は分散が大きな変数を活用しようと動作する。これはすなわち、積極的に変数を活用する方法とそうでない方法であると言える。マーケティング分析で解析対象となるデータは、味や色の好み、価値観、判断基準といった各人で意見が異なるデータである。生理的な欲求など各人で意見が異なる普遍的なデータも存在しているが、普遍的な事柄は解析にかけられるまでもなく「重要」、あるいは、「好まれている」など傾向が明らかであるため、解析対象として重要となるデータは意見が異なるデータである。潜在学習法は分散が大きな変数、すなわち、意見が異なる変数に着目して学習を行う解析手法であるが、この動作メカニズムはマーケティング上の重要データに着目して解析を行うというものであるため、マーケティング分野の解析手法として有効であると言える。潜在学習法の「着目する事柄(変数)を積極的に活用し、結果として重みが削減(変数が絞られる)される」という動作は構造学習と異なっており、繰り返し述べるが、このことはマーケティング上の重要データに着目して解析することであり、潜在学習法はマーケティング分析の解析手法として強みを持っている手法である。

表 6: パラメータ r の変化と各変数の潜在性 (多重共線性データ)

パラメータ r	X1	X2	X3
1.0	1.0000	0.0525	0.2500
1.1	1.0000	0.0340	0.2176
1.2	1.0000	0.0231	0.1895
1.3	1.0000	0.0134	0.1649
1.4	1.0000	0.0049	0.1436
1.5	1.0000	0.0010	0.1250
1.6	1.0000	0.0000	0.1088
1.7	1.0000	0.0000	0.0947
1.8	1.0000	0.0000	0.0825
1.9	1.0000	0.0000	0.0718
2.0	1.0000	0.0000	0.0625

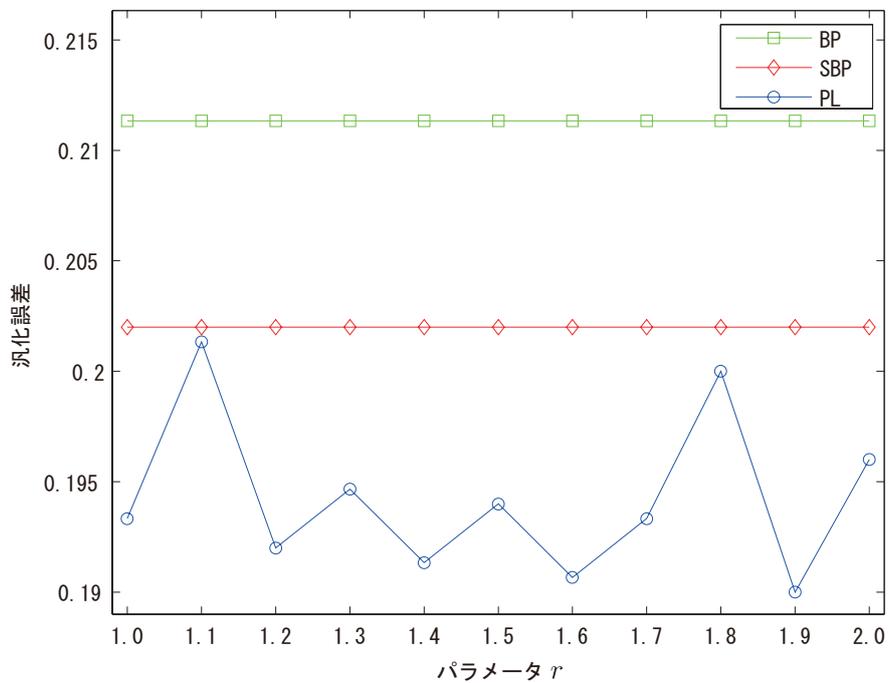


図 50: パラメータ r の変化と汎化能力 (多重共線性データ)

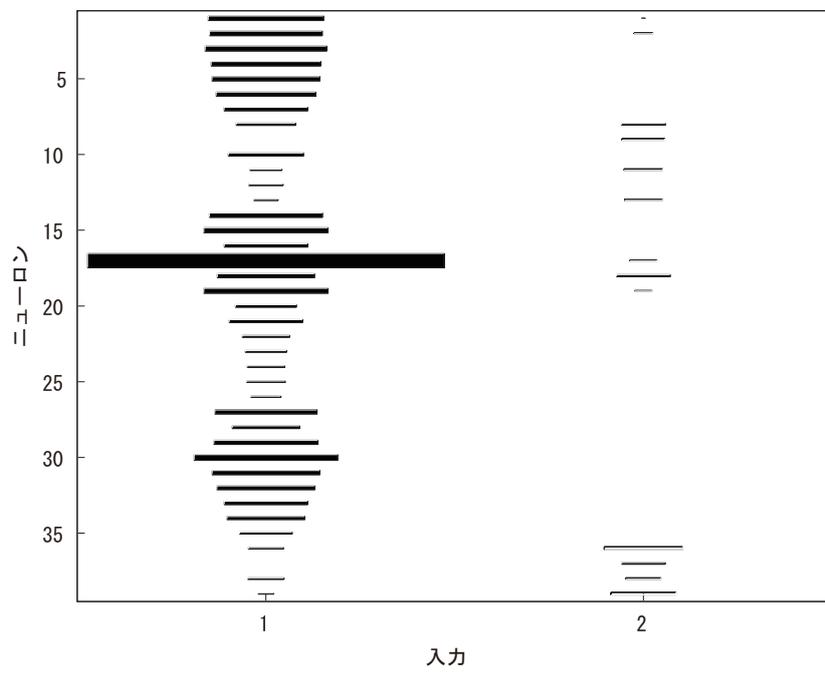


図 51: 潜在学習法の重み (入力層-中間層)

4 消費者来店動機解析

本章ではスーパーマーケットのPOSデータを解析対象として、消費者の来店動機の抽出を試みた解析について述べる。本章で解析対象とするスーパーマーケットではクーポン券による顧客の来店促進施策を実施しているが、来店が途絶えてしまう顧客も存在しこれが課題となっている。この課題に対して解決策となる施策案を提案するべく、「来店が継続する顧客・途絶えてしまう顧客を過去の購買行動から予測するモデル」の作成を試みた。モデルにより継続来店者・非継続来店者を判定し、非継続来店者に絞って追加施策を実施できる環境を整えることを狙っている。解析は「関東地域の1店舗」、「北海道地域の5店舗」、「関東地域5店舗・北海道地域5店舗の合計10店舗」をそれぞれ対象とした。解析の結果「関東地域の1店舗」、「北海道地域の5店舗」を対象とした解析から、概して「食事準備時間帯に生鮮食品を購入する」という行動が来店継続に影響を与えていることが明らかになった（「関東地域5店舗・北海道地域5店舗の合計10店舗」を対象とした解析ではモデル解釈ができなかった）。

4.1 解析概要

本章で述べる解析はスーパーマーケットのID付きPOSデータ (ID-POSデータ) を解析対象とし、顧客のスーパーマーケットへの来店動機の抽出を試みた三つの事例である。

一つめの事例は、関東地域の1店舗を対象としたもの [43]、二つめの事例は北海道地域の5店舗を対象としたもの [56]、そして三つめの事例は関東地域5店舗と北海道地域5店舗の合計10店舗を対象としたものである。これらの解析で使用され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により提供されたID-POSデータを使用している。

解析対象となるスーパーマーケットは継続した来店を顧客に期待しており、その施策として顧客の購買履歴に基づいて作成したクーポン券を各顧客に発券している。このクーポン券は各顧客の購買行動から発券さ

れたものである。基本的にはその顧客が頻繁に購入している商品が安く購入できるようになるという仕組みである。これにより小売店は顧客の継続的な来店を期待し、顧客は必要な商品を安く購入できるという双方にとって利益のある関係構築が期待されている。しかし、クーポン制度を導入しても必ずしも継続的に顧客が来店するとは限らないため、このことが店舗の課題となっている [57]。このようにクーポン発券だけでは顧客の継続来店を期待することは難しく、クーポン以外にも別の対応策が必要である。だが全顧客にクーポン発券と合わせて別施策を実施すると、大変手間でありまたコストも掛かってしまうという問題が発生する。

そこで、手間・コスト面の圧縮という観点からこのことについて考察してみると、問題点が「来店が途切れる顧客」ということから別施策実施対象顧客を絞ることができれば手間・コストの圧縮が可能ということが見えてくる。しかしこの対象顧客を絞る方法で別施策を実施しようとするとなると、新たな問題点が出てきてしまう。それは「どの顧客の来店が途絶えてしまうのか」がわからないという問題である。顧客から来月の来店予定を聞きだせれば問題は解決するがそれは理想論の世界であり、実際に来店予定を顧客に問い合わせると「どうして次回来店を確約しなければならないのか」とかえって来店が途絶えてしまうことになりかねない。

そこで直接的に来店を顧客に問い合わせるのではなく、顧客の過去の来店状況から「来店の有無」を予測する方法を用いることにした。顧客行動データはPOSデータとして蓄積されているので、顧客の過去の購買行動データを入力として「継続来店か非継続来店か」を出力する予測モデルの作成を試みる。

4.2 予測モデル概要

本章で述べる解析で作成を試みるモデルは、過去3ヶ月分の顧客行動データを入力とし、1ヶ月後の来店予測を出力するモデルである（予測モ

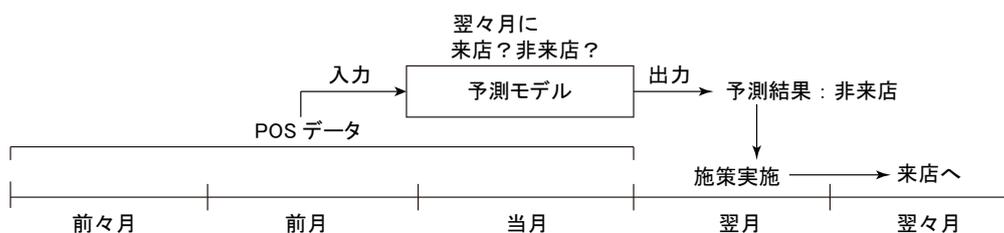


図 52: 非継続来店顧客検知モデル概要図 (文献 [56] より引用)

デルの概要を図 52 に示す)。来店予測を 1 ヶ月先の翌々月にしているのは、この 1 ヶ月間を当該顧客に施策を実施するための期間としているからである。また、入力データの期間を 3 ヶ月にしているのは、季節による購買行動の変化を考慮しているためである。長い期間を入力データとした場合、顧客行動の変化に対応できず予測に対するノイズとなりかねない。そこでビジネスで多用される「四半期」の考え方をを用いて 3 ヶ月を入力期間とした。

本研究で作成するモデルの入力には、スーパーマーケット「全日食チェーン」の POS データを使用する。今回はある年の 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。モデル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 月中に施策を講ずる流れである。

以下、関東地域 1 店舗を対象とする解析は「業績一覧：学術論文 [2] (参考文献 [43])」記載の論文を、北海道地域 5 店舗を対象とする解析は「業績一覧：国内会議 [4] (参考文献 [56])」記載の論文を転載 (構成の入れ替えや一部に加筆・修正は行っている) して述べる (転載について当該学会より許諾を得ている)。

4.3 関東地域 1 店舗を対象とした解析

4.3.1 解析目的

当解析は平成 26 年度データ解析コンペティションにて提供された ID-POS データのうち、関東地域の 1 店舗を対象として非継続来店顧客検知

モデル作成を試みたものである。解析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。モデル作成には潜在学習法を使用する。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。

4.3.2 解析データ

当解析で用いるデータは関東地域のある1店舗の顧客(総数6,998人)を対象とし、12月・1月・2月の3ヶ月の範囲でデータを取得したものである。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ヶ月に渡り表7に示す全変数の値がゼロの顧客(既に来店がない顧客)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²、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(早期停止用)、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。

今回予測に用いたデータでは、全変数がゼロであるデータを4月に来店が無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持つ」というデータだけから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持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

²比率を50:50にしたのは来店がある顧客群に学習結果が引きずられることを防止するためである。

継続来店」と最終出力することを想定している。

当解析では非継続来店顧客を検出するために、表 7 に示す合計 90 個から成る変数を用意した³。これらは五つの変数ジャンル、3ヶ月分の変数であり、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、例えば変数番号 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 7 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 (変数は各月ごと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。

4.3.3 結果と考察

4.3.3.1 非継続来店顧客検知モデル

図 53 は非継続来店顧客検知モデルの汎化能力 (正解率) を示したものである。当解析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上で予測を行っており、図 53 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更させながら予測を行った際の予測区間での平均値である。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、

³データの最頻値は 0 であり、サンプルによっては大きな値を持つという状態であるため、解析時にはデータを対数関数的手法により標準化して使用した。これは x' を標準化後の値とすると、 $x' = \ln(x - \min(x) + 1)$ で求める手法である。

表 7: 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)

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」、潜在性を使用しない潜在学習法と同義である「自己組織化マップを誤差逆伝播法の事前学習に使用する方法、SBP」、そして「潜在学習法、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。当解析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 [58]。また SVM [59] は、分類精度がニューラルネットワークと同等あるいはそれ以上とされており [60]、潜在学習法と

の精度比較を行う目的で実施した。当解析ではソフトマージン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。

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 8 に示す。また、各層のニューロンの個数であるが、SBP 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、SBP の BP 部、PL の学習第 2 段階の中間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⁴ 算出した。

図 53 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、SBP、PL の順で正解率が大きくなり、潜在学習法では $r = 0.3$ 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 (SBP)、そしてさらに潜在性を導入して予測する (PL) ことで、高精度の予測ができることが示された。また、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、潜在学習法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は表 9 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

表 8: 各ニューラルネットワークの構成

		BP		SBP	PL
				BP	学習第 2 段階
伝達関数	中間層	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

⁴SOM Toolbox Version2.0 beta [61] のデフォルト設定。

表 9: モデル別混合行列

n = 608		予測値									
		BP		SBP		PL		Logi Reg		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		197.7	106.4	205.5	98.6	273.5	30.6	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)
		176.2	127.7	133.2	170.7	70.1	233.8	165.4	143.4	31.7	277.1
		(0.2898)	(0.2100)	(0.2191)	(0.2808)	(0.1153)	(0.3845)	(0.2720)	(0.2359)	(0.0521)	(0.4558)

表 10: 各モデルの FPR と FNR

評価尺度	BP	SBP	PL	Logi Reg	SVM
FPR	0.5798	0.4383	0.2307	0.5356	0.1027
FNR	0.3499	0.3242	0.1006	0.2340	0.6033

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 9 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、各モデル化手法共に 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 9 より、最も非継続来店予測を的中 (TP) させるモデルは潜在学習法であること、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。

今回この混合行列に加え、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 10 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分かった。

「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 9 からは「TP と TN の割合が多く、FN の割合が少ないモデル」、表 10 からは「FNR が小さいモデル」、そして、図 53 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言え、これらを総合的に判断して本研究では、潜在学習法が顧客の非継続来店可能性予測には最適な手法であると判断できた。

4.3.3.2 潜在学習法の潜在性の活用

潜在学習法による値の予測には学習第 1 段階で得られた潜在性が活用されている。ここでは潜在積について確認し、モデルがどのような潜在性を活用しているのかを確認する。

図 54 は $r = 0.3$ の際に潜在学習法の学習第 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、データから潜在学習法の学習第 1 段階が獲得した知識を意味している。

図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用いて奥行情報を持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 54 では色を頼りにデータ分類の境界線を判断し、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右端の数値にそれを示している。

今回の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、

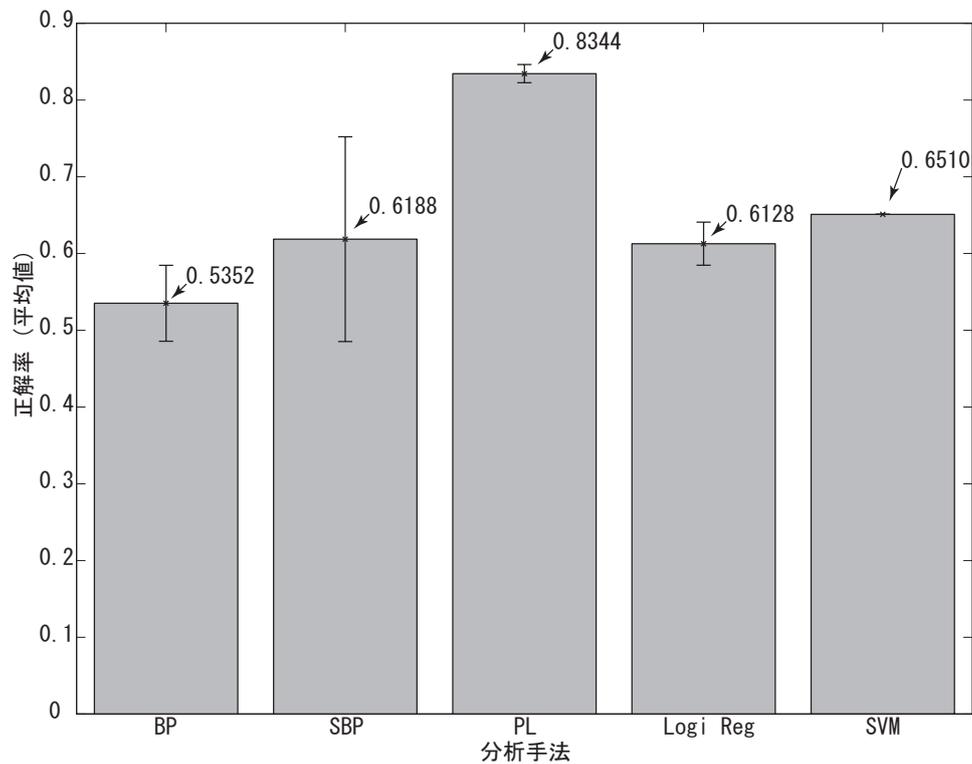


図 53: 各非継続来店顧客検知モデルの正解率

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分かった。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言える。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。潜在学習法の学習第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。今回、潜在学習法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、潜在学習法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。

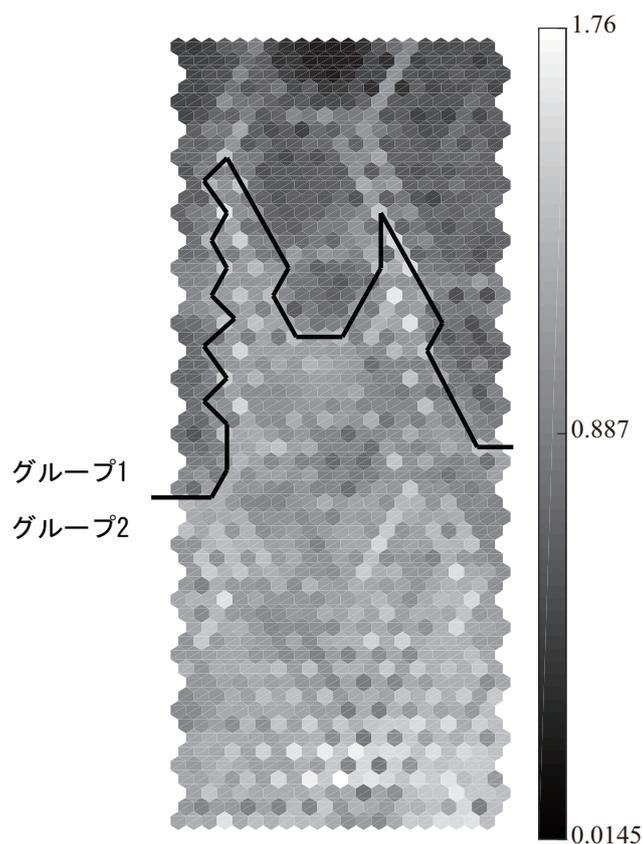


図 54: データから獲得した知識 (潜在学習法、学習第 1 段階での学習結果 ($r = 0.3$))

4.3.3.3 潜在学習法を用いた消費者行動の解釈

図 55 (a) は当解析でデータとして用いた変数の潜在性を示したものである。図 55 (a) 中の x 軸は変数番号を表したものであり表 7 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) が最も潜在性が高い、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 11)、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、潜在性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 位までの潜在性を確認するとあまり差が無い結果であり、どの月であっても重要であると言える。

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 55 (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 55 (b) より、POS に関する変数は各月で同様の潜在性を持っており、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 12)、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、共に大分類 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分かった。

図 55 (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 13)、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分かった。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。

今回得られた重要度のうち、POS 大分類と来店時間に関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ヶ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ヶ月分の合計から算出) の来店回数を占めていることが分かった。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出来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程、継続来店顧客である傾向があることが分かり、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ヶ月分の合計から算出) となっており、おおよそ

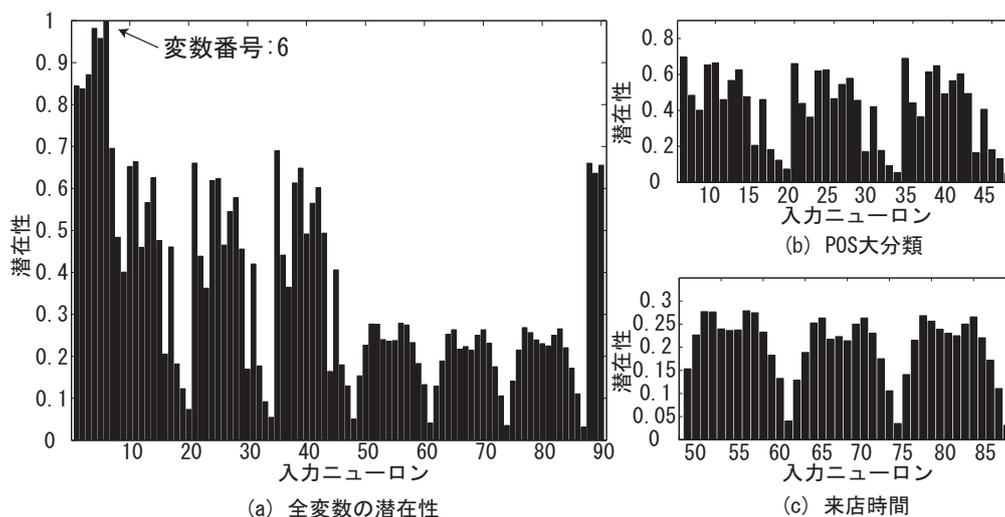


図 55: 潜在学習法による重要変数

表 11: 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: 潜在学習法)

順位	クーポン使用率		総購入数量		総来店回数	
	月	潜在性	月	潜在性	月	潜在性
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

80 : 20 の関係であることが分かった。

4.3.3.4 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 56 (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 53 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 7 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

表 12: 各月の重要 POS 大分類 (上位 3 位: 潜在学習法)

順位	12月		1月		2月	
	POS 大分類	潜在性	POS 大分類	潜在性	POS 大分類	潜在性
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 13: 各月の重要来店時間 (上位 3 位 : 潜在学習法)

順位	12月		1月		2月	
	来店時間	潜在性	来店時間	潜在性	来店時間	潜在性
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

わち重要変数であることが見て取れる。

重要度を変数ジャンルごとに確認すると (表 14)、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他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 56 (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 56 (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) が他の POS 大分類の変数よりも大きなオッズ比を持っていることが分かる。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 15)、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分かるものの、他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 56 (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 であることが分かった。続いて各月ごとに上位 3 位を見ていくと (表 16)、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言える。ただし、これでは全時間帯が重要であるという意味になり、解釈をする意味を持たない結果と言える。

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分かった。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言える。

当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、今回の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53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うが、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で、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言える。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言え、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。

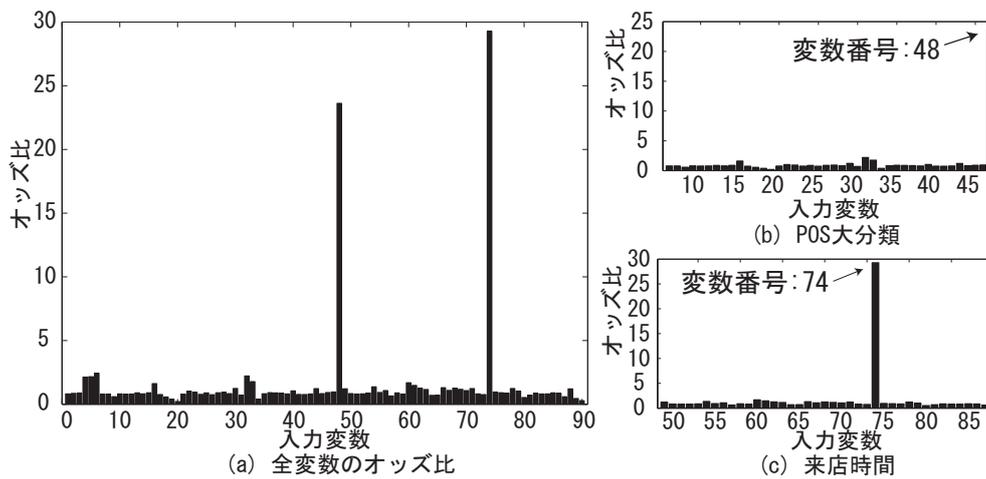


図 56: ロジスティック回帰分析による重要変数

表 14: 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 15: 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 16: 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

4.3.4 まとめ

当解析では小売店のPOSデータを対象とし、判定月より2ヶ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。モデル化には潜在学習法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、潜在学習法の持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であることが分かった。当解析で抽出された消費者行動についてであるが、これは潜在学習法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潜在学習法が適していると判断できた。

また、当解析では90個の変数を使用した。これは多くの変数より特徴を抽出することで予測精度の向上を狙ったことである。解析に潜在学習法を用いた場合は問題は発生しなかったが、ロジスティック回帰分析による解析では計算が終了しないといった問題が発生し、予測精度や解釈性という面以外でも提案手法の有効性が確認できた。

今回使用したデータは、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。今後は他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。

4.4 北海道地域5店舗を対象とした解析

4.4.1 解析目的

当解析は平成26年度データ解析コンペティションにて提供されたID-POSデータのうち、北海道地区5店舗のデータを対象として非継続来店顧客検知モデル作成を試みたものである。[43]の解析では関東地域1店舗のデータを用いて「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」傾向にあることが得られたが、当解析では地域が異なっても同様の事柄が言えるのかどうかを調査する。

4.4.2 データ

当解析で「継続来店顧客予測モデル」を作成するにあたり、解析対象小売店の五つの店舗から得られた総顧客人数91670人から成るデータを作成した。このデータは、ある年の12月、その翌年の1月・2月の3ヶ月分のデータを使用し、4月の来店予測を行うものである。当解析において作成する予測モデルは図52と同様のモデルである。

変数として使用した項目は「クーポン使用率」、「POS大分類」、「総購入数量」、「来店時間」、「来店頻度」の五つのジャンルであり、これを各3ヶ月分用いた。各変数は「クーポン使用率」は百分率、「POS大分類」、「総購入数量」、「来店時間」、「来店頻度」は頻度で値を保持している。

モデル作成前に、このデータについて評価を行った。その結果、POS大分類の大分類14・15において、全サンプルでこれらの変数の値がゼロ、または、少数のサンプルのみで値が出現するということが分かった。よってこれらの変数を削除し、最終的にモデル作成に使用した変数は表17に示す総数87個の変数となった⁵。各変数であるが「クーポン使用率」はその月に発券されたクーポンをどのくらい使用しているか、「POS大分類」は購入商品のジャンル、「総購入数」はその月に合計何個の商品を購入したのか、「来店時間」は来店した時間帯（例えば変数番号49番の「12月

⁵データの最頻値は0であり、サンプルによっては大きな値を持つという状態であるため、解析時にはデータを対数関数的手法により標準化して使用した。これは x' を標準化後の値とすると、 $x' = \ln(x - \min(x) + 1)$ で求める手法である。

表 17: 使用変数一覧

No.	変数名	No.	変数名	No.	変数名	No.	変数名	No.	変数名
1	12月クーポン使用率	21	1月POS大分類04	41	2月POS大分類10	61	1月時間9	81	2月時間17
2	1月クーポン使用率	22	1月POS大分類05	42	2月POS大分類11	62	1月時間10	82	2月時間18
3	2月クーポン使用率	23	1月POS大分類06	43	2月POS大分類12	63	1月時間11	83	2月時間19
4	12月POS大分類01	24	1月POS大分類07	44	2月POS大分類13	64	1月時間12	84	2月時間20
5	12月POS大分類02	25	1月POS大分類08	45	2月POS大分類16	65	1月時間13	85	12月来店頻度
6	12月POS大分類03	26	1月POS大分類09	46	12月総購入数量	66	1月時間14	86	1月来店頻度
7	12月POS大分類04	27	1月POS大分類10	47	1月総購入数量	67	1月時間15	87	2月来店頻度
8	12月POS大分類05	28	1月POS大分類11	48	2月総購入数量	68	1月時間16		
9	12月POS大分類06	29	1月POS大分類12	49	12月時間9	69	1月時間17		
10	12月POS大分類07	30	1月POS大分類13	50	12月時間10	70	1月時間18		
11	12月POS大分類08	31	1月POS大分類16	51	12月時間11	71	1月時間19		
12	12月POS大分類09	32	2月POS大分類01	52	12月時間12	72	1月時間20		
13	12月POS大分類10	33	2月POS大分類02	53	12月時間13	73	2月時間9		
14	12月POS大分類11	34	2月POS大分類03	54	12月時間14	74	2月時間10		
15	12月POS大分類12	35	2月POS大分類04	55	12月時間15	75	2月時間11		
16	12月POS大分類13	36	2月POS大分類05	56	12月時間16	76	2月時間12		
17	12月POS大分類16	37	2月POS大分類06	57	12月時間17	77	2月時間13		
18	1月POS大分類01	38	2月POS大分類07	58	12月時間18	78	2月時間14		
19	1月POS大分類02	39	2月POS大分類08	59	12月時間19	79	2月時間15		
20	1月POS大分類03	40	2月POS大分類09	60	12月時間20	80	2月時間16		

時間9」は12月の9時台の来店を意味している)を、「来店頻度」はその月の来店回数をそれぞれ表している。

また、変数の調査に加えてサンプルの調査も行った。データ作成時点で91670件であったデータは、データ作成期間の3ヶ月間で既に来店が無い顧客などを除き、また、4月の来店と非来店の比率を50対50に調整⁶した結果、サンプル数36294件のデータとなった。

さらに当解析では、このデータを三つのデータ区間に分割して使用した。一つめの区間はモデル作成区間であり、36294件からデータ総数の70%にあたる25406件のサンプルから構成されている。二つめはモデルの学習具合を調整する為のデータ区間であり、データ総数の15%にあたる5444件のサンプルから構成されている。そして三つめは、作成されたモデルの精度を確認する為のデータ区間であり、データ総数の15%にあたる5444件のサンプルから構成されている。この区間分けは上記の比率になるように、データ全体からサンプルをランダムに選択して行った。

4.4.3 モデル化手法

当解析のモデル化において予測モデルは高精度化を狙い、潜在学習法を用いて作成した。本研究と同様にPOSデータの解析をニューラルネッ

⁶比率を50:50にしたのは比率の多い方の顧客群に学習結果が引きずられることを防止するためである。

トワーク (誤差逆伝播法) を用いて行ったものにセールスプロモーション (SP) と売上との関係を学習させ、購買行動のモデル作成を試みた西尾らの研究 [62] がある。この研究では BP に入力データと予測対象データとの関係を学習することを期待している。また、BP 以外の手法では、自己組織化マップ (SOM) を用いて顧客の来店習慣を分類した渡辺らの研究 [63] がある。この研究では SOM で得られた分類結果をさらに別のニューラルネットワーク以外の手法の入力として用いている。ニューラルネットワークの手法が BP であれ、SOM であれ、ニューラルネットワークが学習した内容は「知識」である。このことと上記先行研究でのニューラルネットワークの使用方法をふまえ、本研究では分類と予測の両機能を有している新しいニューラルネットワーク「潜在学習法 (PL)」を使用する。両機能を使用することで、分類により知識を獲得し、その得た結果を予測で活用するということが期待できる。

4.4.4 結果と考察

4.4.4.1 来店予測モデル

図 57 は作成したモデルの精度について示した図である。ここで示している精度は、各データ区間の比率は 4.4.2 で示した比率そのままに、使用するデータサンプルをランダムに変更させて作成した 10 個のモデルから得られた予測精度の平均値である。予測にはモデルにとって未知のデータとなるモデル精度確認用区間のデータを用いているため、モデルの汎化能力が明らかになる。また、ここでいう精度とは、ターゲットの実測値と予測値の誤りの割合である。従って、全ての予測を誤ると 1、全ての予測を的中させると 0 の値を取る。

モデルは精度比較のため、一般的なニューラルネットワークである「BP により作成したモデル」、「潜在性を使用しない潜在学習法と同義となる、SOM で学習した結果を BP で使用するモデル (SBP)」、「PL」、そして本研究のような 2 値の予測時に統計的手法として一般的に使用される [58]、「ロジスティック回帰分析 (Logi Reg)」を用いて、合計 4 種類のモデルを

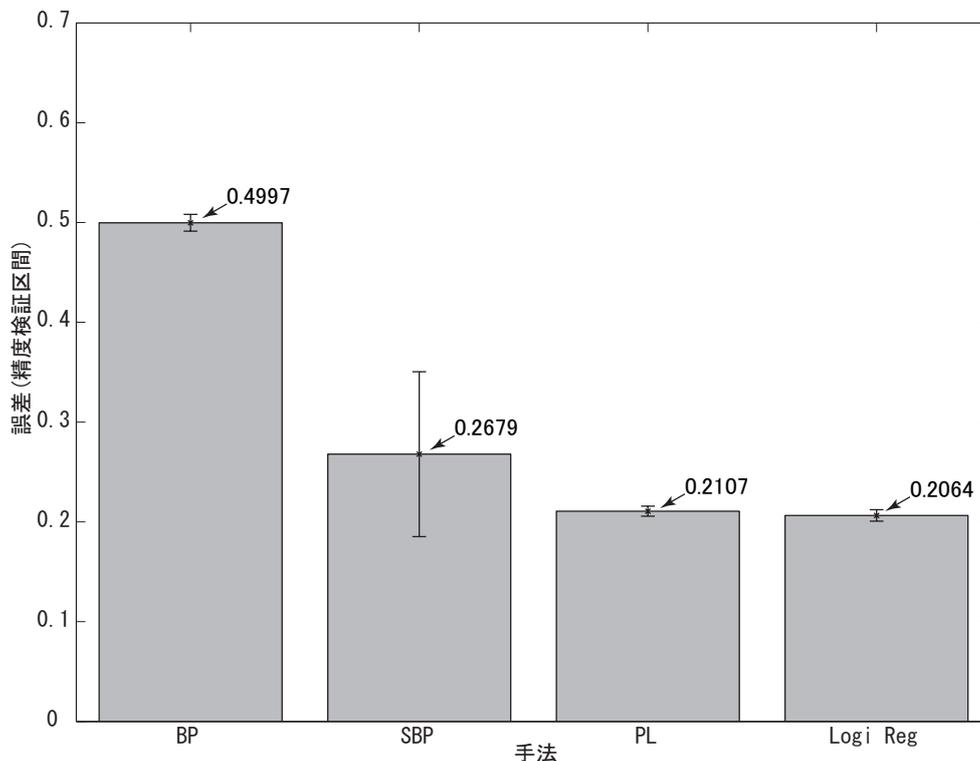


図 57: 作成したモデルの精度

作成した。

各モデルの結果を見ると、精度の悪い順に BP、SBP、PL、Logi Reg という結果であった。これらのうち、BP、SBP、PL はニューラルネットワーク手法なので、これらのみで誤差について考察してみると、与えられたデータから直接的に学習するよりも (BP)、あらかじめデータより知識を得て、それを学習に活用する方が良い精度が得られ (SBP)、そして更に潜在性を使用することでより良い結果を得られる (PL) と解釈できる。また、SBP と PL の差は、誤差のみならず標準偏差も小さくなっており (棒グラフ中に描画されているバーが標準偏差を表している)、PL を使用することで安定したモデルが作成できることが分かった。以上のように、ニューラルネットワークを用いた場合、顧客の来店予測モデル作成には潜在学習法が最も有効であることが分かった。

一方、ロジスティック回帰分析であるが、今回作成した 4 種類のモデルの中で最も高精度なモデルであった。また、標準偏差から確認できるように、安定したモデルであることが分かった。ただし、PL と Logi Reg の誤差の差は 0.0043 で殆ど差がなく、また、両モデルの平均予測誤差に

ついて有意水準 5%で t 検定を行った結果、有意な差は見られなかった ($t=1.7811, df=18, n.s.$)。

このように、予測精度という観点からは潜在学習法によるモデルとロジスティック回帰分析によるモデルの違いについて考察することは難しく、次々節にて消費者行動の解釈という観点で、潜在学習法によるモデルとロジスティック回帰分析によるモデルについて比較を行い、それぞれのモデルについて有用性を確認して行く。

4.4.4.2 潜在学習法が獲得した知識について

消費者行動の解釈を行う前に、潜在学習法が知識獲得段階でどのような知識を獲得しそれが予測に活かされているのかを確認する。

図 58 は知識獲得段階の競合層の出力を示している。知識獲得段階は SOM が基になっているので、ここで示されているのはデータの分類結果である。図 58 (a) は PL による分類結果を示してあり、(b) は SOM による分類結果である。SOM による結果を記載しているのは、手法の違いによる分類結果比較のためである。この図は 2次元の平面の図であるが、使用されている色によって奥行が表現されており、3次元の情報を持っている。奥行のレベルは同系統の色で同程度となっており、寒色系が多い図 58 から考察すると、暖色系はデータの分類結果を区切る境界線と見ることが出来る。

このことより、図 58 (a) における境界線を判断すると、黒色の実線で表されている箇所が、境界線であり、潜在学習法により顧客行動データは二つのグループに分類されると判断できる。一方、図 58 (b) の SOM による分類では、図下部に境界線があると判断できる。

今回、図 58 (a)、(b) 共に、境界線より上部を「グループ 1」、下部を「グループ 2」と命名した。各分類について境界線より考察すると(表 18 に分類結果を示す)、全部で 36294 件あるデータが潜在学習法(図 58 (a))では、ほぼ半分分割された結果であった。一方、自己組織化マップ(図 58 (b))では、偏った分類結果となっていることが分かった。各グループに属するデータ数の調査に加えて、各グループに属するターゲットにつ

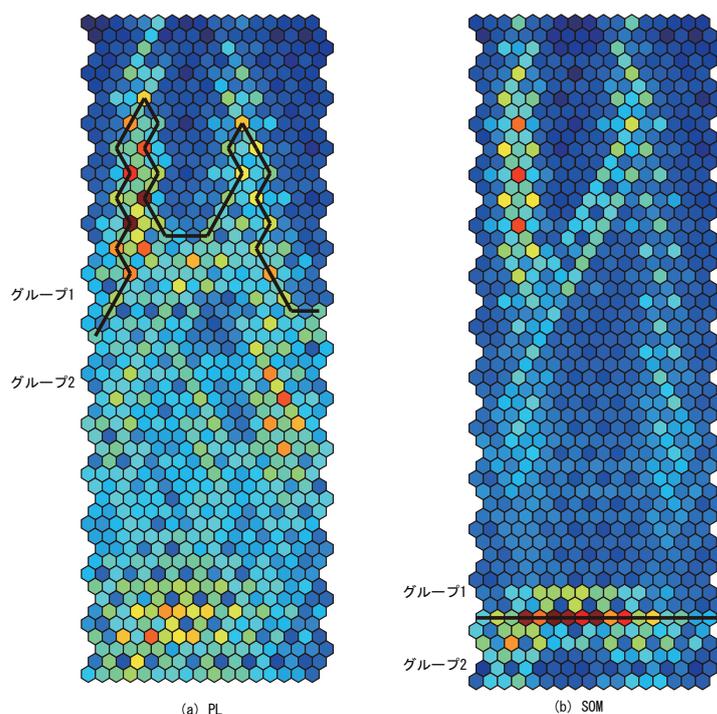


図 58: 獲得した知識

表 18: データ分類結果 (各グループ中のデータ個数)

	潜在学習法		自己組織化マップ	
	グループ1	グループ2	グループ1	グループ2
個数	18008	18286	32177	4117
比率	0.49617	0.50383	0.886565	0.113435

いて調べたところ (表 19 に結果を示す)、潜在学習法 (図 58 (a)) ではグループ1に非来店顧客が14225件、グループ2には3922件含まれており、自己組織化マップ (図 58 (b)) ではグループ1に非継続来店顧客が18000件、グループ2には147件含まれていた。

以上の事柄を総合的に考察すると、潜在学習法はデータを約50:50の比率で2分割し、自己組織化マップはデータを約90:10の比率で2分割した。また、潜在学習法は分割されて生成されたグループ1中にターゲットが1(非継続来店顧客)が約80%、グループ2中にターゲットが0(継続来店顧客)が約80%含まれており、非継続来店顧客のグループと継続来店顧客のグループに分類されていた。

一方、自己組織化マップではグループ1にほぼ100%、非継続来店顧客が含まれており、グループ2に約20%の継続来店顧客が含まれていた。

表 19: データ分類結果 (各グループ中のターゲット別データ個数)

	潜在学習法		自己組織化マップ	
	グループ 1	グループ 2	グループ 1	グループ 2
ターゲット 1				
個数	14225	3922	18000	147
比率	0.783876	0.216124	0.991899	0.008101
ターゲット 0				
個数	3783	14364	14177	3970
比率	0.208464	0.791536	0.781231	0.218769

しかしこの結果は表 18 にて示されている様にデータの分割比率が偏っているので参考にはならない値と判断した。

今回潜在学習法で表 19 の分類結果が得られたということ、知識獲得段階の学習が教師なし学習であることより考察すると、潜在学習法はデータから非継続来店顧客と継続来店顧客の違いを学習したということが言える。また、この知識を獲得するためには自己組織化マップではなく、潜在学習法を使用する必要があることも分かり、潜在学習法の有効性も確認できた。

4.4.4.3 潜在学習法による消費者行動解釈

ここでは、潜在学習法により得られた変数の潜在性から消費者行動の解釈を行っていく。図 59 は入力変数の潜在性を示している。この図を見ると、入力変数番号 1 から 3、4 から 45、46 から 48、49 から 84、85 から 87 の区間において、似た潜在性の値が三つずつ繰り返しているのが見える。これはそれぞれ、12月、1月、2月の同一ジャンルの変数であり、すなわち、各月で変数の重要度は同じであると言え、また、各変数は各変数個別に解釈するのではなく、4.4.2 で示した、「クーポン使用率」、「POS 大分類」、「総購入数量」、「来店時間」、「来店頻度」のジャンル毎に解釈可能であることを示している。

これら五つの変数ジャンルを重要度で並べ替えると、「総購入数」、「クーポン使用率」、「POS 大分類」、「来店頻度」、「来店時間」の順となった。このことより、総じて「どれだけ」、「何を」、「いつ」という三つの購買

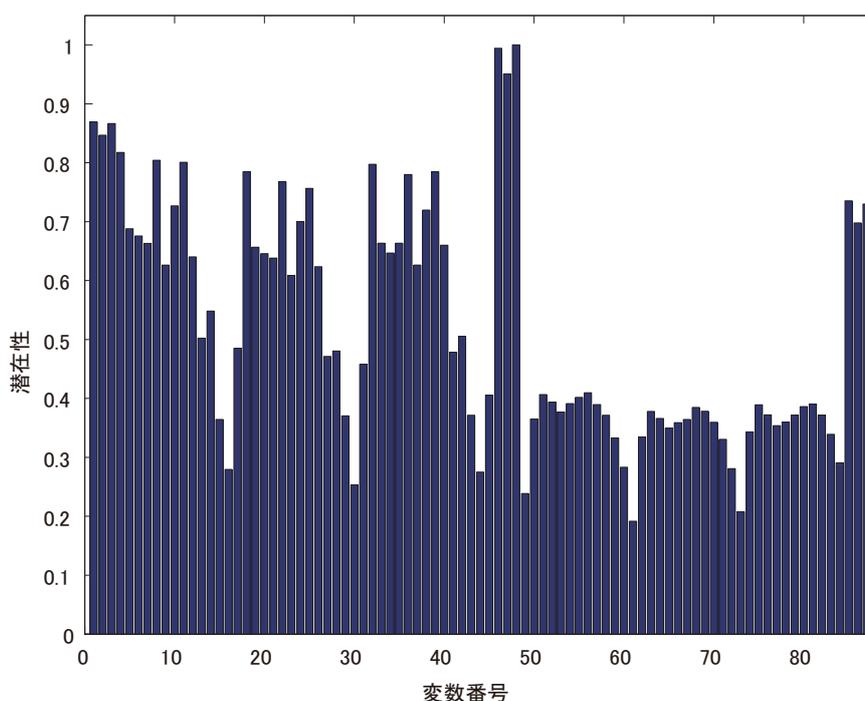


図 59: 入力変数の潜在性

表 20: POS 大分類の重要度上位 3 位 (潜在学習法)

月	重要度		
	1 位	2 位	3 位
12 月	青果	加工食品	一般食品
1 月	青果	加工食品	一般食品
2 月	青果	一般食品	加工食品

行動で来店継続顧客か、非継続顧客かが分かれていると言えることが分かった。ここで変数ジャンルのうち、月情報以外にも詳細な情報を持っている「POS 大分類」と「来店時間」を詳細に確認する。表 20 に POS 大分類での重要度上位 3 位が月毎に示されている。これによると、3ヶ月に共通して「青果」が重要商品であることが示されている。また、重要度 2 位以下では 12 月・1 月と 2 月において「加工食品」と「一般食品」の順位が入れ替わっているものの、やはり共通して重要商品であることが示されている。このことより、各月での POS 大分類に違いはなく、「青果」、「加工食品」、「一般食品」が重要な商品アイテムであると判断した。

続いて、来店時間を確認する。表 21 に来店時間の月毎の重要度上位 3 位が示されている。これを見ると、12 月・1 月・2 月において重要度 1 位の来店時間は 16 時台、17 時台が出現していた。この時間帯は夕食準備

表 21: 来店時間の重要度上位 3 位 (潜在学習法)

月	重要度		
	1 位	2 位	3 位
12 月	16 時台	11 時台	15 時台
1 月	16 時台	17 時台	11 時台
2 月	17 時台	11 時台	16 時台

時間帯と考えることができるので、夕食準備の買物であると判断できる。また、重要度 2 位と重要度 3 位を総じて考察すると、2 位は昼食準備時間帯、3 位は夕食準備時間帯と判断でき、総合的に来店時間は 11 時台と 15 時台から 17 時台の時間帯が重要な時間帯であることが明らかになった。

今回抽出に成功した重要変数から消費者について解釈すると、「昼食・夕食準備時間帯に青果を基本購入商品とし、1ヶ月当たりの購入数量が多い購買行動をしている」顧客が、重要な顧客であるという解釈になった。今回得られたこの重要顧客の行動自体は特に目新しい発見ではないが、重要な事柄である。なぜなら、今回の解析対象である小売店にとっての重要顧客が、「食事準備時間帯に、それ程日持ちのしない青果を中心的に購入している顧客」であるということが抽出されたからである。4.4.1 で述べたが、この小売店は顧客の継続来店を期待してクーポンを発券している。このことより、小売店の期待通りに施策「クーポン発券」が機能していること、「購入 → クーポン発券 → 購入」と来店継続サイクルが働いていることが確認できたと判断できる。

それでは、どのような顧客の来店が途絶えてしまうのかということであるが、重要顧客とは異なる購買行動をとる顧客、例えば「食事準備時間帯に青果などの食材を購入しない顧客」ということになり、これは、一人暮らしや店舗近隣で働いており、食事に弁当を購入している顧客などが推測される。毎食を弁当にする場合、販売されている弁当の種類にも限りがあるので、どこかのタイミングで飽きてしまい、新しい種類の弁当を求めて他の商店に移って行ってしまふことが考えられる。また、他の考えられる非継続来店顧客としては「引っ越し」をした顧客がある。しかし、仮説「弁当」、「引っ越し」を検証するには、店舗近隣のオフィス

表 22: POS 大分類の重要度上位 3 位 (ロジスティック回帰分析)

月	重要度		
	1 位	2 位	3 位
12 月	米	たばこ	一般食品
1 月	その他	加工食品	たばこ
2 月	その他	菓子	雑貨

数や転出転入人口などを調査する必要があり、手元にあるデータからは検証できず、推測の域を出ないため、これらの検証は今後の課題である。

4.4.4.4 ロジスティック回帰分析による消費者行動解釈

前項で行ったのと同様に、ロジスティック回帰分析による消費者行動の解釈を実施する。図 60 は入力変数のオッズ比を示したもので、変数の重要度を示している。この図から判断すると、オッズ比は変数毎に異なる値を示しており、ポテンシャルラーニングの結果とは異なり、各変数ジャンルによる傾向というものも見えてこない。だが、ここでは潜在学習法と解釈基準を統一するために、五つの変数ジャンルを使用し解釈を進めていくことにする。

図 60 のオッズ比より、ロジスティック回帰分析では重要な変数は、「来店時間」、「POS 大分類」、「クーポン使用率」、「総購入数」、「来店頻度」の順であることが分かった。これを総じて解釈すると、「いつ」、「何を」、「どれだけ」という購買行動の違いが来店継続顧客と来店非継続顧客の差ということになる。ただし、重要度 1 位に「来店時間」、重要度 5 位に「来店頻度」が存在しているので、「いつ」に関しては明確に述べることができない。続いて「POS 大分類」と「来店時間」について重要度を調べていく。表 22 は POS 大分類の各月の重要度上位 3 位が示されている。この表からは各月で最も重要な商品アイテムについて把握することが可能である。しかし、各月には統一性が見られない。このため、変数ジャンル「POS 大分類」としての統一解釈はできない結果であった。

来店時間に関しては表 23 に重要度が示されている。この表も POS 大分類と同様に、各月の統一性が見られない結果を示している。全体的に

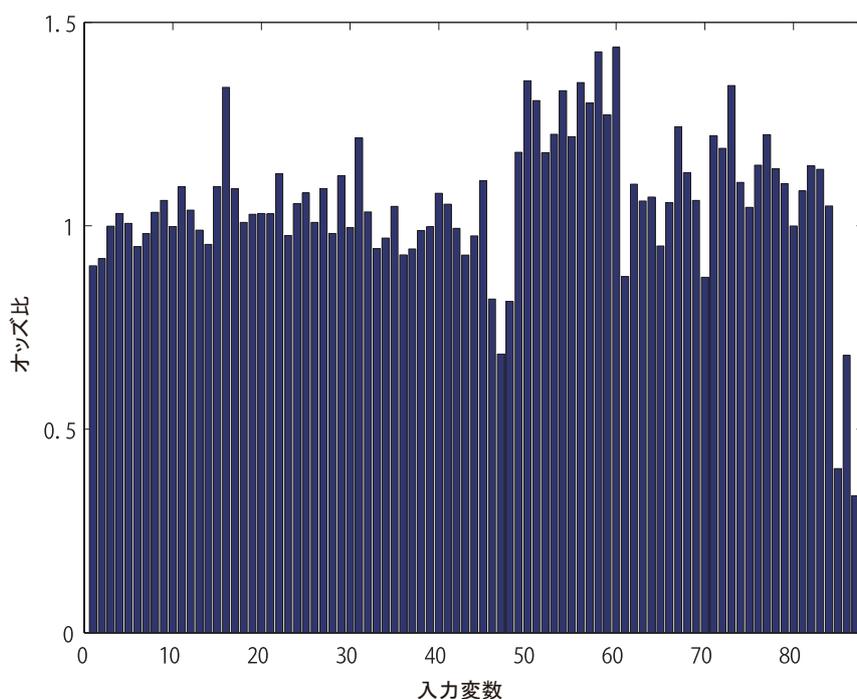


図 60: 入力変数のオッズ比

表 23: 来店時間の重要度上位 3 位 (ロジスティック回帰分析)

月	重要度		
	1 位	2 位	3 位
12 月	20 時台	18 時台	10 時台
1 月	15 時台	19 時台	20 時台
2 月	9 時台	13 時台	12 時台

解釈をするならば、9 時台・10 時台が出現しているので、これを午前中時間帯、12 時台・13 時台・15 時台を日中時間帯、18 時台・19 時台・20 時台を夜間時間帯とし、これらの時間帯が重要であると解釈できる。しかし、これは変数「来店時間」のほぼ全時間が重要であるということであり、重要来店時間の抽出にはならない。よって、ロジスティック回帰分析では変数ジャンル「来店時間」としての統一解釈は不可能であると判断する。

これらのことより、本研究で使用したデータをロジスティック回帰分析にて解釈することはできないと結論付けた。よって、4.4.4.1 で持ち越された、潜在学習法とロジスティック回帰分析の本データに対する有効性は、潜在学習法に軍配が上がる結果となった。また、今回、ロジスティック回帰分析が一般的にニューラルネットワークで言われているように、高精

度であるが解釈が難しいという結果となったことは意外な結果であった。

4.4.5 まとめ

本研究では小売店のPOSデータを入力データとする、「継続来店顧客予測モデル」の作成と、消費者行動の解釈を潜在学習法と称する新しいニューラルネットワークと統計手法であるロジスティック回帰分析を用いて試みた。

研究の結果、ニューラルネットワークで平均予測誤差0.2107という高精度のモデルを作成することに成功した。また、精度比較のためにロジスティック回帰分析を使用してモデルを作成したが、こちらは平均予測誤差0.2064であり、ニューラルネットワークによるモデルと同様に高精度なモデルが作成できた。予測精度を観点とすると、どちらもほぼ同じ予測精度であり、また、統計的に有意な差も見られず、モデル作成手法による違いは見られない結果となった。

しかし、観点を消費者行動の解釈に移すと違いが表れた。ニューラルネットワークによるモデルでは、「昼食・夕食準備時間帯に青果・加工食品・一般食品を購入する顧客」が重要な顧客であるということが抽出されたことに対し、ロジスティック回帰モデルでは、このような解釈を行うことができないという結果となった。

一般的にニューラルネットワークによるモデルはブラックボックス問題により解釈が難しいとされているため、本研究では解釈を可能とするために潜在学習法を使用した。本研究で「ニューラルネットワークでのみ解釈が可能」という結果を得たことよりその試みは期待通りであったと結論付ける。

従来手法では抽出できない事柄が存在するということは、従来の方法では明らかにならなかった事柄を潜在学習法を使用することで明らかにできる可能性を示唆していると言え、今後はこの潜在学習法をさまざまなデータに対して適用し、従来手法との解釈の違いが発生するかなど、マーケティング諸問題の解決手法として活用できる手法であるかどうかを研

究していく。

また、本研究で解析対象とした小売店のデータは、データ提供元となる平成26年度データ解析コンペティションから、関東地区5店舗と北海道地区5店舗の計10店舗分のデータが提供されている。今回は北海道店舗の5店舗だけを使用して解析を行ったが、次回は全データを対象に解析を行い、10店舗で共通して使用できる継続来店・非継続来店顧客の予測モデルを作成すると共に、地区による顧客行動の差異は存在するのかなど、更なる消費者行動を明らかにして行きたい。

4.5 全店舗を対象とした解析

4.5.1 解析目的

当解析は、平成26年度データ解析コンペティションにより提供された全店舗（関東地域5店舗、北海道地域5店舗）のID-POSデータを解析したものである。

これまで実施してきた関東地域1店舗を対象とした解析 [43] からは「青果、加工食品、惣菜を購入している顧客が重要」なこと、北海道地域5店舗を対象とした解析 [56] からは「青果、加工食品、一般食品を購入している顧客が重要」なことが抽出された。どちらの解析でも共通した「食料品」の購入顧客が重要であるとの結果であるが、関東地域に関しては未解析店舗が4店舗存在しており、上記二つの解析をもって「全日食チェーン全体での重要顧客」を論ずることは時期尚早である。そこで当解析で全店舗を対象に解析を実施する。

4.5.2 解析データ

当解析では二つのデータを解析に用いた。一つはモデル作成用データ（12月、1月、2月から成る4月来店予測用）で、49218件のデータ（来店あり顧客と来店なし顧客の比率は50:50⁷）である。もう一つはモデル精度確認用データ（1月、2月、3月から成る5月来店予測用）で、50481件のデータである。使用データの変数については表24に示している。

来店有無予測モデルは、モデル作成用データの全サンプルを用いて作成され、モデル精度確認用データでモデルの汎化能力を調査される流れとなっている。

4.5.3 解析手法

モデルは深層学習を用いて作成した。ネットワークは出力層にソフトマックス関数を持つ4層構造のネットワークである。深層学習は積層自

⁷比率を50:50にしたのは比率の多い方の顧客群に学習結果が引きずられることを防止するためである。

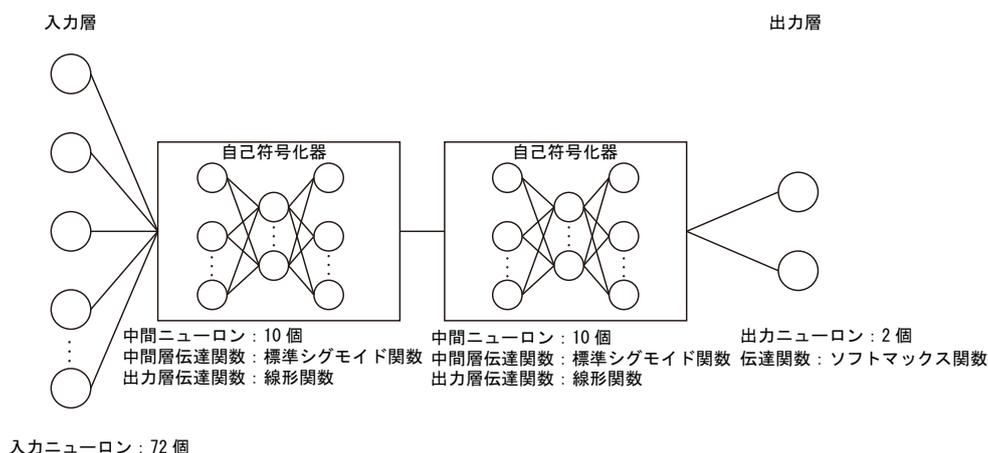


図 61: 使用したネットワーク概要図

自己符号化器を使用して事前学習を行った。

ネットワークの構成は、一つの入力層、二つの自己符号化器、一つの出
力層を持った、合計4層構造をしている(概要を図61に示す)。また、入
力層のニューロンは72個、出力層のニューロンは2個であるが、これは
表24に示す72変数からなるデータを入力データとしていることと、出
力(来店あり・来店なし)を2値で示しているからである。なお、自己符
号化器の中間ニューロン数や中間層伝達関数は解析に用いた数値計算言
語「MATLAB」の標準設定を使用している。

また、精度比較として深層学習(DL)から一つの自己符号化器を除い
たネットワーク(自己符号化器による事前学習を行う誤差逆伝播法(以後
「事前学習併用誤差逆伝播法」と称する, ABP)、そして事前学習なしの誤
差逆伝播法(以後「誤差逆伝播法」と称する, BP)も用いた。比較のため
に用いた誤差逆伝播法は、中間層に10個のニューロン、中間層の伝達関
数に双曲線正接関数、出力層の伝達関数にソフトマックス関数を用いて
いる。これはMATLABの標準設定である。

4.5.4 結果と考察

4.5.4.1 深層学習

深層学習によるモデル作成の結果、誤判定率が0.2206のモデルが作成
できた。このモデルの混合行列を図62(a)に示す。この図より、モデル
は来店あり顧客を33.0%、来店なし顧客を45.0%検出していることが分か

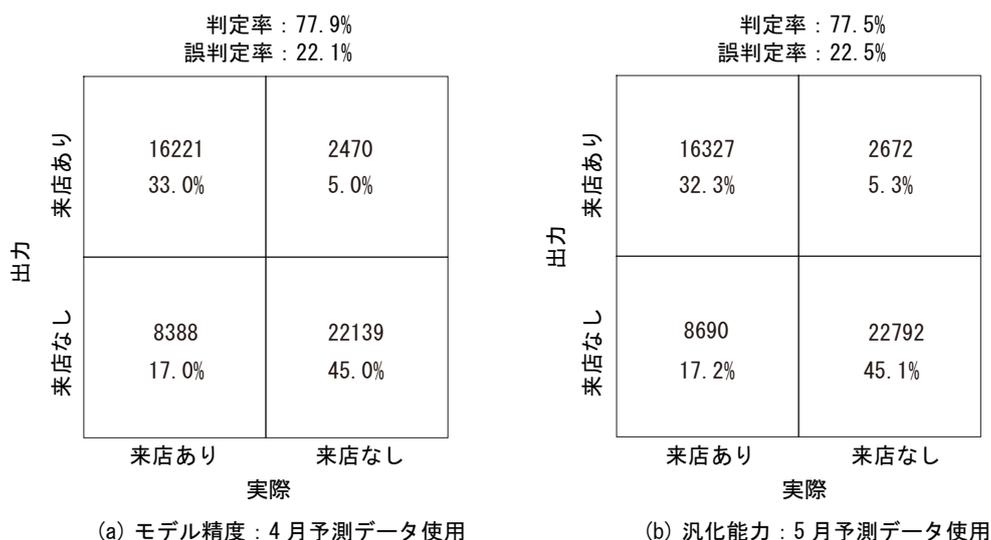


図 62: 深層学習によるモデル化とモデル汎化能力

る。データ中の来店あり・来店なしの比率は 50:50 であるので、来店なし顧客の検出が来店あり顧客の検出と比べて精度が高いことがわかる。また、誤判定であるが、「実際には来店がある顧客」を誤って「来店なし」と判定してしまう傾向が図から見て取れる。

このモデルに汎化能力確認用データ (5月予測用データ) を投入し、汎化能力を確認したものが図 62 (b) である。誤判定率は 0.2251 であった。図 62 (b) より、判定結果、誤判定結果共に作成したモデルと大きく異なることはなく、同程度の結果であることが見て取れる。

4.5.4.2 事前学習併用誤差逆伝播法

続いて、事前学習併用誤差逆伝播法を用いてモデル化した結果について確認を行う。事前学習併用誤差逆伝播法によるモデルの誤判定率は 0.2206 であり、混合行列 (図 63 (a) に示す) を確認すると、モデルは来店あり顧客を 33.1%、来店なし顧客を 44.8% 検出していることが分かる。また、誤判定であるが、「実際には来店がある顧客」を誤って「来店なし」と判定してしまう傾向が図から見て取れる。

このモデルに汎化能力確認用データ (5月予測用データ) を投入し、汎化能力を確認したものが図 63 (b) である。誤判定率は 0.2242 であった。深層学習の際と同様に図 63 (b) より、判定結果と誤判定結果は共に、作成したモデルと大きく異なることはなく、同程度の結果であることが見

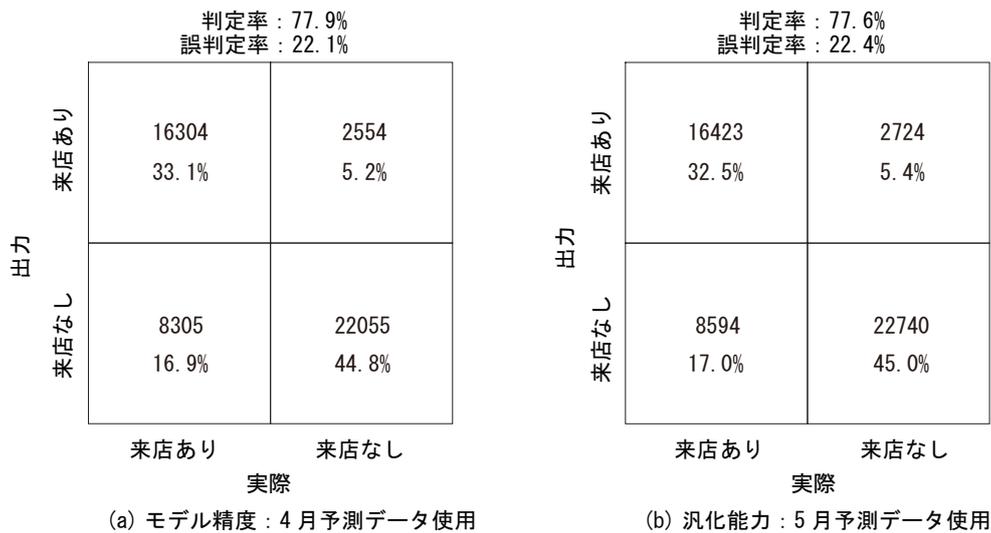


図 63: 事前学習併用誤差逆伝播法によるモデル化とモデル汎化能力

て取れる。

4.5.4.3 誤差逆伝播法

モデル精度の比較のために、事前学習を行わない誤差逆伝播法を用いてモデル作成を行った結果を示す。誤差逆伝播法によるモデル作成の結果、誤判定率が0.2199のモデルが作成できた。混合行列(図64(a)に示す)を確認すると、モデルは来店あり顧客を34.2%、来店なし顧客を43.8%検出していることが分かる。また、誤判定であるが、「実際には来店がある顧客」を誤って「来店なし」と判定してしまう傾向が図から見て取れる。

このモデルに汎化能力確認用データ(5月予測用データ)を投入し、汎化能力を確認したものが図64(b)である。誤判定率は0.2234であった。深層学習の際と同様に図64(b)より、判定結果と誤判定結果は共に、作成したモデルと大きく異なることはなく、同程度の結果であることが見て取れる。

4.5.4.4 重要変数解釈

ここでは入力変数のうち、重要な変数を抽出する目的で各ネットワークの重みを確認する。図65、図66、図67がそれぞれのネットワークの重みを示している。図中の四角形は重みの大きさを、黒色は符号がプラス、白抜きは符号がマイナスであることを意味している。

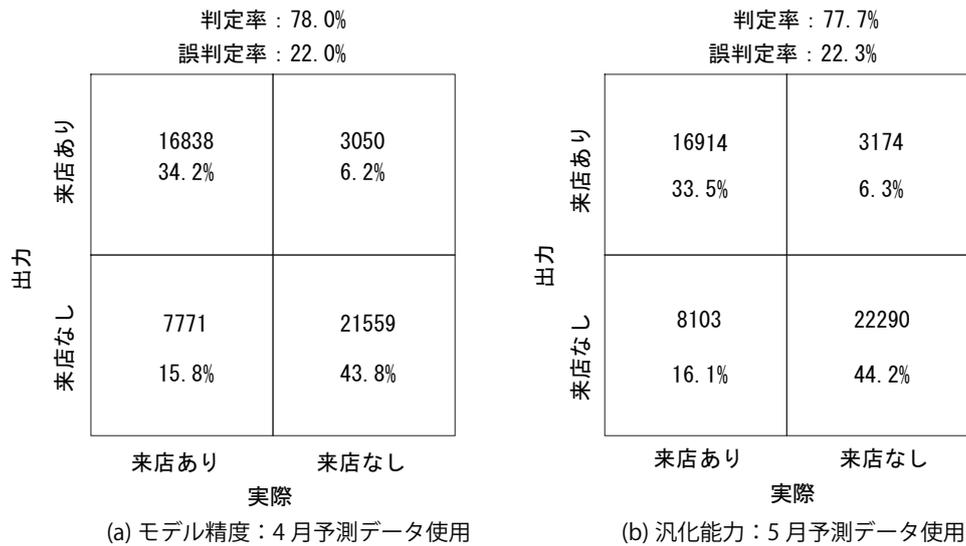


図 64: 誤差逆伝播法によるモデル化とモデル汎化能力

図 65 に示す深層学習の重みであるが、図 65 (a) を確認すると、変数番号 36 付近を境にして、それより変数番号が前の重みと後の重みで傾向が変わっていることが確認できる。前後を比較すると後の方は粗密な印象を受ける。変数は変数番号 37 から時間に関する変数に変わっており (表 24 参照)、これより、POS 大分類に関する変数の方が、時間に関する変数よりも影響力を持っていると判断ができる。続いて 1 層目の自己符号化器の重みであるが (図 65 (b)) この図からは各ニューロンがさまざまに値を持っていることが分かる。2 層目の自己符号化器の重みでも (図 65 (c)) 同様である。入力から出力までは一連の流れであるので、入力層-1 層目の自己符号化 (図 65 (a)) で影響力を持っていた変数 (重み) がどのように出力層へ伝わるのかを追跡する必要がある。

しかし、入力層-1 層目の自己符号化器 (図 65 (a)) から 1 層目の自己符号化器-2 層目の自己符号化器 (図 65 (b)) へ伝わる段階で既に変数番号 36 付近前後の傾向は別として、各ニューロンがどのように作用しているのか分からない状態である。よって、深層学習により重要変数の解釈を行うことは、POS 大分類と時間に関する変数で影響力が異なるという点以外では困難であると結論付けた。

事前学習併用誤差逆伝播法の重みは図 66 に示している。図 66 (a) より、入力層-自己符号化器では深層学習の結果と同様に (図 65 (a)) POS

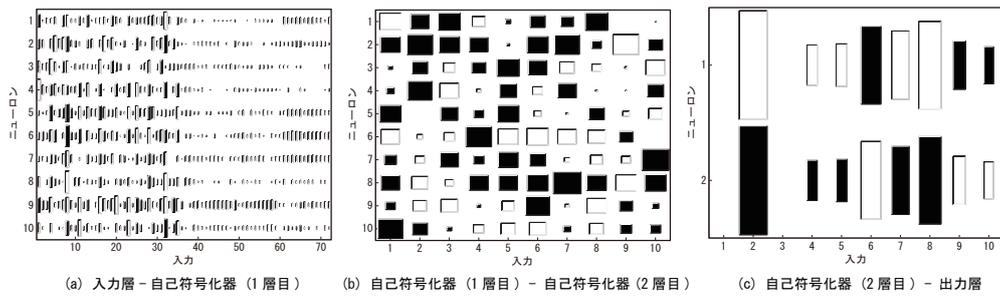


図 65: 深層学習の重み

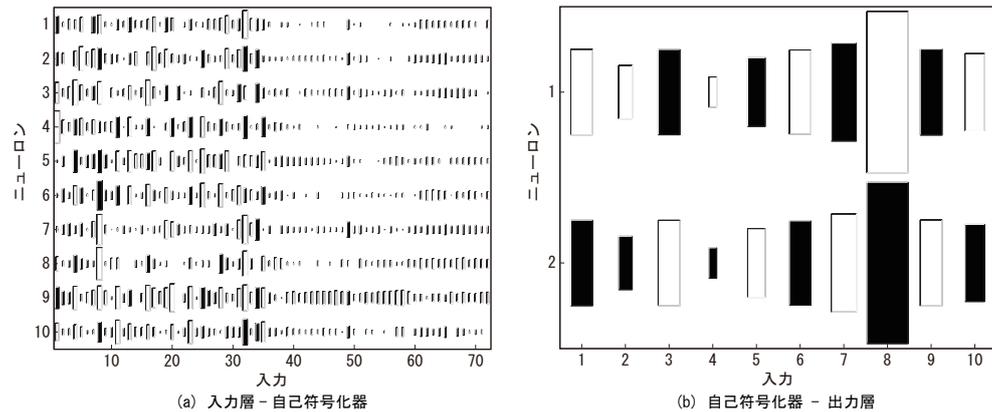


図 66: 事前学習併用誤差逆伝播法の重み

大分類の方が時間に関する変数よりも大きい重みを持っており、影響力に変数ジャンルで違いがあるといえる結果であった。一方、自己符号化器-出力層 (図 66 (b)) は重みの大きさ、符号も各出力ニューロンで異なっている結果であった。各層共に活動しているニューロンが多く、よって特定のニューロンについて言及することはできない結果であった。

誤差逆伝播法の重みは図 67 に示している。誤差逆伝播法では深層学習・事前学習併用誤差逆伝播法と異なり、入力層-中間層の重みから (図 67 (a)) 変数ジャンルの影響力の違いについて確認することはできなかった。中間層-出力層の結果は (図 67 (b)) 事前学習併用誤差逆伝播法と同様に重みの大きさ、符号も各出力ニューロンで異なっている結果であった。誤差逆伝播法は事前学習併用誤差逆伝播法以上に各層で活動しているニューロンが多く、やはり特定のニューロンのみには言及することができない結果であった。

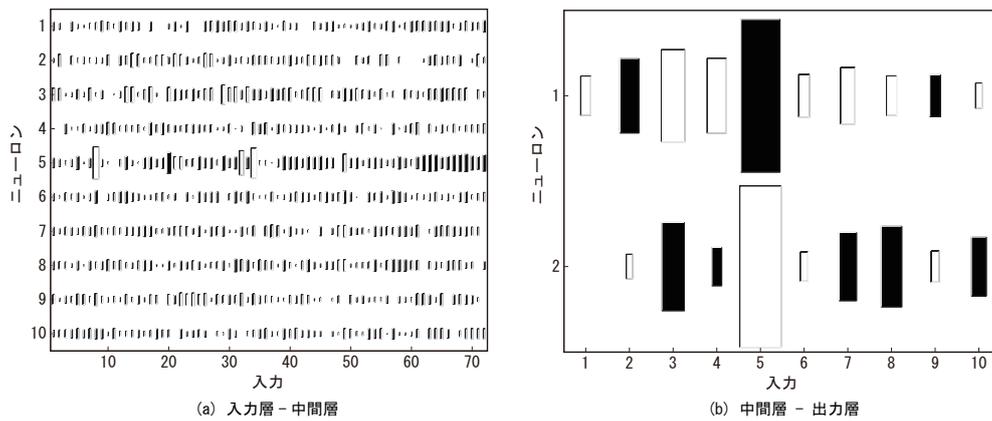


図 67: 誤差逆伝播法の重み

4.5.5 まとめ

深層学習によるモデル、事前学習併用誤差逆伝播法によるモデル、誤差逆伝播法によるモデルを比較すると、全体的な精度は誤差逆伝播法によるモデルが最も精度がよい結果となった。2番目に精度がよいモデルは深層学習と事前学習併用誤差逆伝播法によるものでこれら両者に精度の違いはなかった。

しかし、モデルの特性を図 62 (a)、図 63 (a) そして図 64 (a) を用いて比較すると、深層学習の来店なし検出率が 45.0%、事前学習併用誤差逆伝播法の来店なし検出率が 44.8%、通常 BP の来店なし検出率が 43.8%であること、「実際には来店があるのに来店なしと誤判定してしまう率」が深層学習で 17.0%、事前学習併用誤差逆伝播法で 16.9%、通常 BP で 15.8%であり、モデルの傾向はそれぞれ異なっている。この誤判定率を、「実際には来店がないのに来店ありと誤判定してしまう率」と比べることにより深層学習によるモデルは、来店なしの検出量が多いモデル、通常 BP によるモデルは、来店ありの検出量が多いモデルであると言える。

「来店あり顧客を来店なしと誤判定してしまう」ことと、「来店なし顧客を来店ありと誤判定してしまうこと」を比べると、前者は来店予定者に過剰に来店を促すこととなり、後者は来店が途絶えてしまう顧客に何の対策も行えないことを意味する。過剰な来店促進活動は顧客が辟易し、マイナスの効果を生むとも考えられるものの、来なくなってしまう顧客に何もできない方が実務面から考えると損失であると考えられる。この

表 24: 使用変数一覧

変数番号	変数名	変数番号	変数名
1	前々月 POS 大分類一	37	前々月時間九
2	前々月 POS 大分類二	38	前々月時間十
3	前々月 POS 大分類三	39	前々月時間十一
4	前々月 POS 大分類四	40	前々月時間十二
5	前々月 POS 大分類五	41	前々月時間十三
6	前々月 POS 大分類六	42	前々月時間十四
7	前々月 POS 大分類七	43	前々月時間十五
8	前々月 POS 大分類八	44	前々月時間十六
9	前々月 POS 大分類九	45	前々月時間十七
10	前々月 POS 大分類十	46	前々月時間十八
11	前々月 POS 大分類十一	47	前々月時間十九
12	前々月 POS 大分類十三	48	前々月時間二十
13	前月 POS 大分類一	49	前月時間九
14	前月 POS 大分類二	50	前月時間十
15	前月 POS 大分類三	51	前月時間十一
16	前月 POS 大分類四	52	前月時間十二
17	前月 POS 大分類五	53	前月時間十三
18	前月 POS 大分類六	54	前月時間十四
19	前月 POS 大分類七	55	前月時間十五
20	前月 POS 大分類八	56	前月時間十六
21	前月 POS 大分類九	57	前月時間十七
22	前月 POS 大分類十	58	前月時間十八
23	前月 POS 大分類十一	59	前月時間十九
24	前月 POS 大分類十三	60	前月時間二十
25	当月 POS 大分類一	61	当月時間九
26	当月 POS 大分類二	62	当月時間十
27	当月 POS 大分類三	63	当月時間十一
28	当月 POS 大分類四	64	当月時間十二
29	当月 POS 大分類五	65	当月時間十三
30	当月 POS 大分類六	66	当月時間十四
31	当月 POS 大分類七	67	当月時間十五
32	当月 POS 大分類八	68	当月時間十六
33	当月 POS 大分類九	69	当月時間十七
34	当月 POS 大分類十	70	当月時間十八
35	当月 POS 大分類十一	71	当月時間十九
36	当月 POS 大分類十三	72	当月時間二十

ことより、深層学習によるモデルの方が予測特性の面で誤差逆伝播法によるモデルよりも店舗に有益であると結論付けた。しかし、全体精度では誤差逆伝播法の方がよく、層が深いという深層学習の効果が活かされてはいないと言える結果となった。これは事前学習併用誤差逆伝播法による学習結果と比較しても導かれる。深層学習と事前学習併用誤差逆伝

播法は全体精度もモデル特性もほぼ同じである。よって深層化することで精度が必ずしも向上するわけではないという結果であった。

また重要変数の解釈であるが、一般的にニューラルネットワークでは重みの解釈は困難であるとされており、深層学習であってもこの傾向は変わらないと言える結果であった。加えて、深層学習は通常のニューラルネットワークよりも層が多いため、変数の影響力の可読性は通常のものよりも低くなることも言える結果であった。なぜなら、どの層を解釈すればいいのかという問題が発生するからである。多層であるため確認を要する重みが増えることになるが、個々の層でネットワークが何を学習しているのか可読性が低いために解釈が困難であり、そのため多層化のメリットが活かせていないのである。画像を対象とした深層学習であれば、各層で得られた結果を可視化することで、各層でどのような特徴が抽出されているのか把握ができるが、データを対象とした深層学習の場合は可視化しても解釈に結びつかないことがあると言える結果であった。

5 履修科目決定行動解析

本章では消費者行動モデルに基づいたアンケートより得たデータを解析し、大学生が授業履修決定時に重要視する事柄の抽出を試みた解析について述べる。アンケートは「精緻化見込みモデル」に基づいており、学生がシラバスや授業ガイダンスより提供される情報を処理している「ルート」を明らかにするものである。アンケートから作成される解析データは高精度のデータ処理を期待してニューラルネットワークにより解析される。解析の結果、「中心ルート」と「周辺ルート」の各ルートを経て授業情報を処理する学生の割合は約3対7であり、このルートの違いには設問「以前からその先生の授業を受けてみたかった」が強く影響していることが明らかになった。

5.1 解析目的

当解析は「大学生が履修科目を決定する際に重要視している事柄」を明らかにすることを目的とした解析である [64]。

科目履修前に学生が授業について公式に内容を知る機会は、大学が事前配布するシラバスからでしかない。学生はこの情報を基に履修科目を仮決定し、そして初回の授業で行われるガイダンスを受講して履修の最終決定を行うが、学生は授業に関する情報をどのように処理して理解あるいは説得されているのだろうか。当解析はこのことを明らかにすることを狙っている。解析データは消費者行動研究のモデルである「精緻化見込みモデル (Elaboration Likelihood Model, ELM)」に基づいたアンケートにより収集された。以下解析について「業績一覧：国内会議 [7] (参考文献 [64])」記載の論文を転載 (一部に加筆・修正は行っている) して述べる (転載について当該学会より許諾を得ている)。

5.2 精緻化見込みモデル

精緻化見込みモデル (Elaboration Likelihood Model, ELM) は Petty と Cacioppo により提唱された説得コミュニケーションに関するモデルであり [65], [66], [67], [7]、情報を受け取った消費者の情報処理ルートを示したものである。情報処理ルートは二つ存在し、一つは「中心ルート」、もう一つは「周辺ルート」と呼ばれるものである。中心ルートは「得た情報を検討し処理するルート」であり、周辺ルートは「得た情報を本質的な部分ではない周辺情報から処理するルート」である。周辺情報とは例えば「製品パッケージのデザイン」や「広告に出演している俳優」、当解析の内容であれば「教員の雰囲気」や「友人と一緒に履修する」などといった本質 (授業内容) 以外のことである。すなわち、中心ルートは授業に関して本質的内容を吟味したうえで履修を判断する学生がたどるルート、周辺ルートは授業の本質ではない情報から履修を判断する学生がたどるルートである。

5.3 消費者行動モデルと学生履修行動

当解析では大学生の履修行動解析に消費者行動のモデルを使用するが、これは「授業は大学が学生という消費者に提供するサービスである」と捉えているからである。

大学で開講されている授業には履修が必須な必修科目と、履修が自由な自由選択科目があるが、このうち自由選択科目の履修科目選択行動において学生は、大学が提供する授業というサービスを消費する消費者と捉えることができる。科目の中には、あまり消費されない授業 (履修者が若干名もしくは履修者なし) も存在し、この履修状態が改善されない場合はその授業は廃止されてしまうこともある。授業の廃止は、学生が今後その科目について学べなくなるという学習機会の損失に繋がるため避けたい事態である。このような事態を避けるため、大学は学生のニーズを捉えて授業を提供する必要がある。大学は教育機関であり営利目的ではないとの声もあろうが、授業の提供と履修、そして廃止の関係から鑑み

ると営利目的企業と同様にマーケティングの考え方を導入する必要がある。Kotler は著書のなかで大学を非営利団体のサービス業であると記しているが [68], [69]、このことも大学が学生ニーズを考慮する必要があること支持するものである。

しかし大学は教育の場であるので営利団体と同様にはニーズに応えることはできないという制約があることもまた事実である。例えば、多くの学生のニーズを純粹に推測すると「容易に単位を取得できること」となるであろうが大学は高等教育機関であり、また大学が持っている「育成したい人材像」に関するポリシー (ディプロマポリシー) に反してしまうことになるのでそれに応えることはできない。

よって、大学が応えることのできるニーズは、学生の「学習したい事柄」と「授業の方法や運営方法」についてということになる。これらのうち「学習したい事柄」については学生から直接的に汲み取り授業を設けるという方法もあるだろうが、教員が近年の社会情勢などから判断し、今またはこれから必要となる知識について授業を開講する方が現実的であるので、応えるべきニーズとまでは言えない。また、教員はその道の専門家であるので、余程のことが無い限り、ここにニーズのミスマッチは起こりえないであろう。そうであるならば大学が応えることができるニーズは「授業の方法や運営方法」についてである。だが、消費者は自分の欲しいものについて自覚していないあるいは具体的に表現できない場合があると言われている [68]。特に学習においてそれはことさらであろう。よって、学習に関して明確な目的と要望がある学生からはニーズを収集し運営方法に反映、それ以外の学生に対しては、授業の内容・運営方法について担当教員の Unique Selling Proposition (USP) などの説明を通して履修を説得させる必要がある。当解析ではこの学生の被説得過程を解析するため、説得コミュニケーションに関する精緻化見込みモデルを適用させた。

当解析は、アンケート調査とデータ解析から構成されている。アンケート調査により学生の履修行動についてデータを収集し、データ解析により履修行動で最も重要な事柄を明らかにするという流れである。このう

ち、アンケート調査に精緻化見込みモデルが用いられている。

5.4 調査方法と解析方法

5.4.1 アンケート調査

履修に至る過程を調査するために合計 29 の問いから成るアンケートを作成・実施した。アンケートは二つの設問群から構成されている。一つは「中心ルートを経て履修を決定するのか・周辺ルートを経て履修を決定するのかを調査する設問群」であり、ELM の枠組みを用い 7 問から構成されている (この 7 問からアンケート回答者の情報処理ルートが明らかになる)。もう一方は、周辺ルートを経た場合の「履修の決め手」について調査する設問群で、こちらは 22 問より構成されている。

調査時期は 2014 年度の春semesterと秋semesterであり、東海大学湘南キャンパスの学部生が対象である。調査を実施した授業は主に情報教育センター開講科目である。合計 17 科目 (春semester 4 科目、秋semester 13 科目) の回答者総数 599 人、有効回答率 86.48% で得られた、518 サンプルを解析対象とした。なお、このアンケート調査は『東海大学「人を対象とする研究」に関する倫理委員会』の承認を得て実施した。

5.4.2 解析方法

アンケート調査により得られた履修過程に関するデータは、学生が重要視している事柄を抽出するためにデータ解析にかけられる。ここでの狙いは、「履修の決め手」に関する 22 問の回答結果を入力変数として、ELM に基づいた 7 設問から得られた履修決定に至るルート (中心ルートか周辺ルートか) を予測し、何がルートの決定に影響を与えているのかを抽出することである。「履修の決め手」に関する設問は本来、周辺ルートを経た際の「決め手」を知るための設問であるが、意思決定はあるルートのみを経て行われているわけではないとされているため [70]、よって、本アンケートではルートを問わず履修の決め手について質問しており、そしてこの設問を変数として用いることで、中心ルートを経た履修決定者に

についても周辺の事柄の何を重要視しているのかを知ることが期待している。

アンケート結果を解析するにあたり、当解析では誤差逆伝播法を用いて解析を行った。ネットワーク構造は入力ニューロンが22個、中間ニューロンが41個、出力ニューロンが2個である。出力ニューロンが2個あるのは、ルートを2値で(中心ルート:01、周辺ルート:10)表しているためである。また、中間ニューロンと出力ニューロンの伝達関数には双曲線正接関数を使用している。ニューラルネットワークによる解析は、パラメータの初期値がランダムに与えられることから解析の度に結果が異なるという問題点があるが、当解析では10回解析を繰り返し平均誤差を求めた。使用データについては、データ全体の70%に当たる362サンプルをモデル構築用、全体の15%に当たる78サンプルをモデル構築時の検証用(早期停止用)、全体の15%に当たる78サンプルをモデル精度のテスト用とし、3分割して使用した。また、10回の繰り返しの過程では、このデータ分割の比率はそのままに解析に使用するサンプルもランダムに変更し、使用サンプルの偏りを回避している。

アンケート結果の解析には前述したとおり誤差逆伝播法を用いるが、解析後のモデル解釈にはインフォメーションロスと呼ばれる手法を用いる[71], [72]。この手法はニューロンが何らかの理由により欠落した状況を想定したものである。インフォメーションロスは、学習が完了したネットワークから入力ニューロンを順次欠落させていき、最も情報が損失するニューロンを探索する。これにより例えば5番目の入力ニューロンを欠落させた際に情報損失が最も大きいとすると、この5番目の入力ニューロンに対応する変数が最も重要な(影響力を持つ)変数であると解釈する手法である。

5.5 結果と考察

5.5.1 アンケート調査結果

アンケート調査の結果、中心ルートを経て履修に至った学生は31.85%、周辺ルートを経た学生は68.15%であり、約3対7の比率である事が明らかになった。

Semester毎の結果を表25に記す。春 Semesterの比率が若干異なるが、これは春 Semesterでのアンケート実施科目数が少ないためである。よって、 Semesterが変わってもルートの比率は大きく変化しないと判断した。

履修行動における中心ルートと周辺ルートは「意欲的な受講生」と「それほどでもない受講生」と言い換えることができる。当解析で得られた学習意欲に関する比率は、教員に有益な情報である。従来は担当教員が受講生の学習意欲に関して感覚的に比率を感じて、講義スタイルを調整していたが、今後はこの割合を基準値として調整することが可能となるからである。

5.5.2 データ解析結果

データ解析の結果であるが、周辺ルートに関する22設問を変数とし、履修に至るルートを誤差逆伝播法により予測した結果、モデル精度テスト用サンプルにおいて平均予測誤差25.00%を得た。得られた予測精度について精度の良否を確認する目的で、同一のデータをロジスティック回帰分析にて予測を行ったところ、平均誤差27.05%を得た。この結果により、誤差逆伝播法の方がデータをよく学習しており、この学習結果を使用してデータ解釈を行うことに問題はないと判断した。予測誤差の詳細は表

表 25: 履修時通過ルート (Semester平均)

Semester	中心ルート (%)	周辺ルート (%)
春・秋 Semester	31.85	68.15
春 Semester	35.53	64.47
秋 Semester	31.22	68.78

解析手法	平均誤差 (%)	最大誤差 (%)	最小誤差 (%)	標準偏差
誤差逆伝播法	25.00	32.05	21.79	2.91
ロジスティック回帰分析	27.05	30.77	23.08	3.05

26 に記す。

5.5.3 変数解釈結果

続いて、予測誤差 21.79% (最小値) を得た誤差逆伝播法の結果に、インフォメーションロスを適用し重要変数を抽出した。その結果、アンケート設問:「以前からその先生の授業を受けてみたかった」の回答結果に対応する変数の情報損失が最も大きく、すなわちこの変数が最も重要な役割を果たしているという結果を得た。

重要変数として抽出された、変数 (アンケート設問:「以前からその先生の授業を受けてみたかった」) がルートの違いに与える影響について解釈を行った。アンケートの集計結果より、中心ルートの学生の多くが、周辺ルートの学生よりもこの問いに対して肯定する回答をしていることが分かった (比率の差の検定の結果、1%水準で有意)。この結果、中心ルートの学生はシラバス情報から履修決定を行うと共に、「以前からその先生の授業を受けてみたかった」という過去に形成された意思を用いて最終的に履修を決定しているという構造が明らかになった。

5.6 まとめ

当解析は、授業履修決定要因の抽出を試みたものである。解析データは精緻化見込みモデルに基づいて作成したアンケートを実施して収集した。収集したデータは誤差逆伝播法により解析され、授業履修判断時に考慮した事柄を入力、情報処理ルートを出力とするモデルが作成された。モデルは平均予測誤差 25.00%で作成され、これはロジスティック回帰分析によるモデル精度 (平均誤差 27.05%) よりも高精度であった。モデルはインフォメーションロスにより、モデル作成に使用したデータ変数の

うちで最も重要な変数を抽出され、その結果、「以前からその先生の授業を受けてみたかった」という設問が重要であることがわかった。この設問は過去に形成された意思について問うた設問であり、履修決定意思はシラバスやガイダンスといった現時点の情報からではなく、それよりも前に形成された意思から影響を受けていることがわかった。

今回の研究では、過去の意思についての調査はできていないため、過去意思の形成経緯、また過去意思の現在の意思決定への影響度合について調査していくことが今後の課題である。

6 重要 Tweet 判断基準抽出

本章は災害時に発信された Tweet を解析対象として、災害時に Tweet を重要と判断する基準の抽出を試みた解析について述べたものである。解析は「重要情報を含む Tweet と含まない Tweet を分類するモデル」を作成し、このモデルを解釈することによって実施する。本章では二つの解析について述べる。一つは重要 Tweet 分類モデル作成について検証を実施したもの、もう一つはモデル作成とモデル解釈を実施したものである。モデル作成には高精度の予測結果を期待してニューラルネットワーク (誤差逆伝播法と潜在学習法) を用いた。解析の結果、Tweet を重要・非重要に分類する高精度なモデルが作成でき、モデル解釈の結果、Tweet を構成する文字数が重要・非重要の分類に影響を与えていることが明らかになった。

6.1 解析概要

当解析は災害発生時の Twitter データ (東日本大震災発生時のデータ) を解析し、重要情報を含む Tweet とそうでない Tweet を分類するモデルの作成を目指したものである。モデルの作成後はモデルについて解釈し、人が Tweet を重要と判断する際に重要視する事柄について抽出を試みる。ここでは二つの解析についてとりあげる。一つめは分類モデルの作成について検証したもの [73]、二つめはモデル作成にくわえて重要情報の抽出を試みたもの [74] である。

6.2 災害と Twitter

近年、Twitter や Facebook といったソーシャルネットワーキングサービス (SNS) が自然災害発生時に重要な役割を果たしている [75] [76]。それは Twitter が従来のメディアとは異なり、現場にいる多数の人物から情報が発信されるからである。

Twitter 登場以前においては、情報を収集する主な手段はテレビやラジオ、Web サイトなどであった (表 27 に各メディアの特性を示す)。テレビはニュース番組、特集番組、速報テロップなどを通じ避難情報や安否情報といった情報を提供してくれる。ラジオも同じく、ニュース番組や特集番組にて情報を提供してくれる。Web サイトの場合は情報を必要なタイミングで入手することができる。しかし、メディアそれぞれには特徴があり、例えばテレビは広範囲に情報を発信することが可能である一方、特定の地域の詳細な情報を入手することが難しい場合も多い。ラジオの場合は地域放送局を聴取すればテレビよりも狭い範囲の情報を入手することができるが、やはりある程度広範囲の情報となってしまう。また両者とも放送のタイミングといった問題もあり、必要な情報が必要な時に入手できるとは限らない。一方、Web サイトは必要とする地域に関する情報を必要なタイミングで入手することが可能である。ただし、情報を発信しているサイトにたどり着くことができないと情報を入手することが出来ない。また、発信している情報が、テレビやラジオよりも局所的な情報であっても限界がある。詳細な状況のリアルタイムな情報発信はその現場に居る人物にしか速報性の面で出来ないのである。

そこで SNS が活用されることになるが、しかし、自然災害発生時には大量の情報が SNS を通して発信されることになるため情報が錯そうし、また、被災者は情報収集だけを行える状況にないため人々の情報収集は容易ではない。くわえて、Twitter は基本的にはコミュニケーションツールであり、災害時の Tweet であっても災害情報が必ず含まれているとは限らない。そのため、例えば被災者が災害関連情報を得ようとする場合、多数の Tweet の中から災害に関する Tweet のみを探し出し、その中から自分に必要な情報を見つけ出す必要がある。

情報の収集には電力が必要である。大規模災害発生時に被災地域で電力が変わりなく供給され続けているということは考えにくく、一時的にしろ停電が生じていると考えるのが自然である。よって、情報収集能力は収集装置 (スマートフォンなど) に搭載されている電池の残電力に影響される。状況によっては必要事項を収集したら一時的に収集装置の電源

を切り、電力の温存に努めなければならない場合もある。従って、長時間にわたって目的の情報を発信している Web サイトや Tweet を見つけるために検索ワードを試行錯誤している余裕はなく、素早く容易に必要なとする情報を入手できる方法が必要である。

そこで Tweet を自動的に分類するモデルについて作成を試みた。Tweet 内容が災害時に役立つ情報であるか・否か (重要か・非重要か) によって分類し、このうち重要情報を含む Tweet 群を情報収集源として検索することで、検索を素早くできることを期待している。また、当解析では Tweet を分類するだけでなく、分類モデルが入力変数のどの項目に着目して分類を行っているのか抽出を試みる。解析で使用する Tweet には、人手により「重要」か「非重要」かのラベルを付与したものを使用した。この Tweet データを用い、分類する際にモデルが着目した変数を解析することで、人間が災害時の状態で重要と判断する情報がどのような事柄であるのか抽出を試みることができる。

当解析と関連する研究に情報通信研究機構が開発した、対災害 SNS 情報分析システム DISAANA (DISAster-information ANAlyzer) [77] の存在があげられる。これは災害発生時に SNS を活用するためのシステムであり、投稿数が膨大となる Tweet を効率良く検索できることを目標としている。この DISAANA は Tweet をリアルタイムに解析し、システムに対して質問が投げかけられると解析した内容に基づいて回答を行うシステムである。Tweet の解析は「文分割」、「形態素解析」、「係り受け解析」、「地名処理」、「モダリティ解析」、「属性判定」、「予報表現抽出」、「拡張パターン抽出」といった処理からなる「言語処理モジュール」にて行われ、質問に対する回答を作成、そして情報収集者が投げかけた質問に対して回答を提供するという流れでシステムが構成されている [78]。DISAANA は自治体にて有効性検証実験が行われるなど高い成果をあげているものである。

しかし、DISAANA には当解析で試みる Tweet が重要かどうかを判断する処理は存在しない。よって、質問内容によっては避難に無益な情報や特定の人物に向けて発信されている情報も出力されてしまう。当解析

表 27: 各メディアの特徴 (文献 [73] より引用)

メディア	情報取得方法	発信される情報の内容	情報取得に必要な事柄
テレビ	受動的	全体的	放送電波へのチューニング
ラジオ	受動的	やや全体的	放送電波へのチューニング
Web サイト	能動的	局所的	検索能力
Twitter	受動的	とても局所的	フォロー

では人間が災害時にどのような情報を重要と判断するのかという事柄の抽出を試みるが、これは Tweet 分類の際のフィルタリングなどに応用できる有益な情報である。このように DISAANA と当解析は SNS からの情報検索を容易にしたいという共通点はあるものの異なる研究である。

以下、モデル作成検証に関する解析は「業績一覧：国内会議 [6] (参考文献 [73])」記載の論文を、モデル作成と重要情報の抽出に関する解析は「業績一覧：学術論文 [1] (参考文献 [74])」記載の論文を転載 (構成の入れ替えや一部加筆・修正および翻訳は行っている) して述べる (転載について当該学会より許諾を得ている)。

6.3 モデル作成概要

モデル作成は Tweet データに対して形態素解析 (形態素解析には「JUMAN」 [79] を用いた) を実施するところから始まる。形態素解析により Tweet データから単語情報を抽出して出現状態を確認、解析用データとする。そしてその解析用データを解析しモデル作成を実施するという流れである。手順について以下に示す。

1. Tweet データを形態素解析にて形態素に分解する。
2. 形態素データから解析用データを作成する。
3. 解析手法を用いて Tweet 内容を「有益」・「無益」に分類するモデルを作成する。
4. 分類結果を考察する。

6.4 Tweet 分類モデル作成検証

6.4.1 解析目的

当解析は Twitter の Tweet データをニューラルネットワークを用いて、災害に関する情報を含む Tweet と、含まない Tweet に分類することを試みる。本章でとりあげる解析は Tweet 分類モデル作成と、重要と判定する際に重要視される事柄の抽出であるが、当解析ではモデル作成に関する部分のみを解析範囲とする。

6.4.2 Tweet データの形態素解析

当解析では東日本大震災時にやり取りされた Tweet から 1000 サンプルをランダムに抽出し、解析対象として使用した。解析の最初の処理として、Tweet データに形態素解析を適用する。

当解析では形態素解析器として「JUMAN」[79] を利用した。JUMAN の辞書には「ドメイン」と「カテゴリ」という項目が用意されており、解析用データの変数はそのドメインとカテゴリの出現頻度を主に用いた。カテゴリとドメインとは言葉についての詳細事項のことである。例えば「コンピュータ」という言葉はカテゴリ「人工物-その他」、ドメイン「科学・技術」というものになる。

6.4.3 解析用データの作成

表 28 と表 29 は当解析で使用したカテゴリとドメインを示したものである。JUMAN に最初から用意されているカテゴリは 1 番から 22 番までの 22 種類、ドメインは 1 番から 13 番までの 13 種類であるが、今回はこれにカテゴリを 3 種類 (23 番から 25 番)、ドメインに 122 種類 (14 番から 135 番) を追加した。追加したものは、どのような Tweet データが情報として有益であるかということを考察した結果である。特に追加ドメインは場所についての事柄であり、情報を活用するには場所情報が必要であろうという仮説に基づいている。実際、場所情報は重要なものであり

表 28: カテゴリ一覧

番号	カテゴリ名	番号	カテゴリ名
1	人	16	場所-機能
2	組織・団体	17	場所-その他
3	動物	18	抽象物
4	植物	19	形・模様
5	動物-部位	20	色
6	植物-部位	21	数量
7	人工物-食べ物	22	時間
8	人工物-衣類	23	ハッシュタグ
9	人工物-乗り物	24	リンク
10	人工物-金銭	25	地名:日本
11	人工物-その他		
12	自然物		
13	場所-施設		
14	場所-施設部位		
15	場所-自然		

Tweet データより場所情報を抽出しようとする研究はいくつも行われている [80], [81], [82]。

解析用のデータはこれら「カテゴリ」と「ドメイン」の出現頻度、そして「ツイートの文字数」の大きく分けて3変数を用いて解析データとした。ただし、大きく分けて3変数であるが、詳細に個数を述べると160変数から成っている。具体的には表28に記載の25個のカテゴリと表29より番号7の「ドメイン無し」を除いた134個のドメイン、そして「Tweetの文字数」の合計160変数である。なお、表29の注釈(※1)にあるように「XXX」は都道府県名を表しているが、番号67から90のドメインにおいては北海道から中部地方までと、近畿地方の一部の都道府県名までとなっており、全都道府県名ではないことをここで断っておく。

6.4.4 Tweet 分類モデル

前項の一連の作業により解析用のデータが作成されたので、このデータを入力データとする Tweet データの分類器作成を試みる。

当解析では分類手法として誤差逆伝播法を用いる。また、解析用デー

表 29: ドメイン一覧

番号	ドメイン名	番号	ドメイン名
1	文化・芸術	16 - 60	交通-鉄道-XXX ※ 1
2	レクリエーション	61 - 65	交通-鉄道-XXX ※ 2
3	スポーツ	66	交通-鉄道線
4	健康・医学	67 - 90	交通-鉄道線-XXX ※ 1
5	家庭・暮らし	91 - 135	展示場-XXX ※ 1
6	料理・食事		
7	ドメイン無し		※ 1 : 都道府県名
8	交通		※ 2 : 地域名
9	教育・学習		
10	科学・技術		
11	ビジネス		
12	メディア		
13	政治		
14	交通-国道		
15	交通-高速道路		

タは、形態素解析を行った 1000tweet から 600tweet をランダムに抜き出し、600 サンプルのデータとした。これは、有益 Tweet と無益 Tweet の個数をほぼ同数にしたためである。「有益」・「無益」の情報は災害に関する Tweet や Tweet 中に具体的な地名や建物名が含まれているなど、避難情報として活用できると判断したものを「有益」とし、目視にて手作業で割り当てた。

ネットワークにデータを入力するにあたり、変数をそれぞれ考察した結果、全ツイートにおいて出現頻度が 0 の変数や出現頻度が極端に少ない変数が存在したのでこれを削除した。この作業により解析用データの変数は表 30 に示す 25 変数となった。

図 68 は今回使用したネットワークの概要図である。ネットワークは 3 層構造で、入力層のニューロンは 25 個、中間ニューロンは 33 個、出力層のニューロンは 2 個を使用した。今回ネットワークから得たい出力は Tweet が「有益」か「無益」を表すラベルであり当解析ではラベルを 2 値で示しているため、出力ニューロンは 2 個である。また、中間ニューロンの伝達関数には双曲線正接関数を、出力ニューロンの伝達関数には標準

表 30: 解析用データに採用した変数の一覧

番号	変数名	番号	変数名
1	人	16	文化・芸術
2	組織・団体	17	家庭・暮らし
3	人工物-乗り物	18	料理・食事
4	人工物-その他	19	交通
5	自然物	20	教育・学習
6	場所-施設	21	ビジネス
7	場所-自然	22	メディア
8	場所-機能	23	政治
9	場所-その他	24	交通-鉄道線-東京都
10	抽象物	25	文字数
11	形・模様		
12	数量		
13	時間		
14	地名:日本		
15	リンク		

シグモイド関数を使用している。なお、このラベルの数は 600 データ中、有益が 294 個、無益が 306 個である。

中間層のニューロン数は、ニューロン数を 1 個から 100 個まで変化させ、またそれぞれの隠れニューロン数で重みの初期値がランダムに異なる 10 個のネットワークを用いて学習させた結果を総合的に判断し 33 個と決定した。

データは有益・無益のデータからそれぞれ 7 割を学習用データ、2 割を検証用データ、1 割をテストデータとして抽出して使用した。各データの個数と有益・無益の内訳は、学習用データが合計 420 個 (有益 : 206 個、無益 : 214 個)、学習中の検証用データが合計 121 個 (有益 : 59 個、無益 : 62 個)、テストデータが合計 59 個 (有益 : 29 個、無益 : 30 個) である。検証用データは過学習を防ぐ目的で採用した早期停止法を行うために利用した。

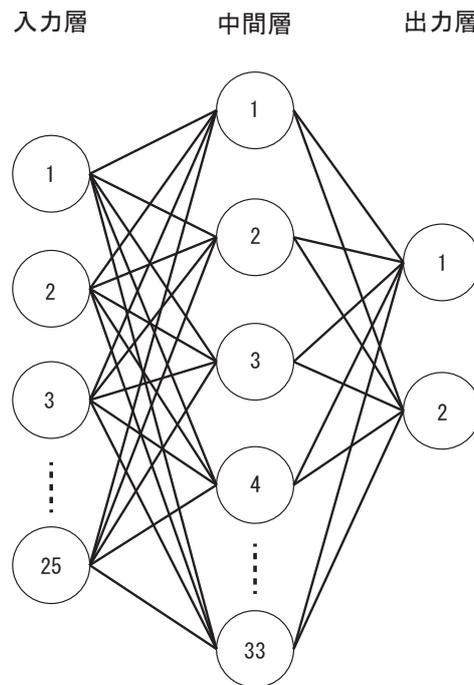


図 68: Tweet データ分類に使用したネットワーク

6.4.5 分類結果

このデータを誤差逆伝播法にて分類した結果、誤判定率は学習データにおいて 22.38(%)、テストデータにおいて 13.56(%) で分類ができた (図 70 「誤差逆伝播法」)。図 69 は隠れニューロン数が 33 個で、重みの初期値がそれぞれ異なる 10 個のネットワークを作成して学習させた際の誤判定率を示している。なお、ここで学習させた 10 パターンの初期重みが異なるネットワークは、前記の隠れニューロン数を決定した際の 10 パターンのネットワークとは違い、33 個の隠れニューロン数での最良の結果を探索する目的で新たに作成したものである。当解析で分類結果として示した誤判定率 (学習データ:22.38(%)、テストデータ:13.56(%)) は図 69 に示す、パターン 10 のネットワークを用いた際に得られたものである。これら 10 パターンの重みの初期値が異なるネットワークを用いた学習における平均誤判定率は、学習データで 27.12(%)、テストデータで 31.69(%) であった。また最大誤判定率と最小誤判定率は学習データで 50.95(%)、15.71(%)、テストデータで 52.54(%)、13.56(%) であった。

ここで誤差逆伝播法による分類精度を考察するために、他の手法を用いてもこのデータを分類し、分類精度を比較することにする。今回はロ

ロジスティック回帰分析を用いた。分類精度は、学習データで18.81(%)、テストデータで38.98(%)であった(図70「ロジスティック回帰分析(変数番号24削除)」。ただし、この分類結果を得るためにはデータを修正する必要があり、結果を単純に比較することはできない。修正した理由は誤差逆伝播法で使ったデータではロジスティック回帰分析にて処理できなかったためである。回帰の係数を求めるために反復計算を行っており、当初その回数を100回として計算したが反復回数内で結果を得ることができなかった。この反復回数を10000回まで増加させてみたがそれでも収束することはなく、このデータを用いてロジスティック回帰分析にて分類することはできないと結論づけた。反復回数100回の時点での結果を見たところ、変数24のオッズが異常に大きい値であった。そこで変数24を削除してみたところ100回以内の反復で結果を得ることができたので、この結果をロジスティック回帰分析での解析結果とすることにした。

このようにして二つの手法によりTweetの分類ができたわけであるが、これら得られた誤判定率より判断して、誤差逆伝播法による分類結果は高い精度でありTweetデータの分類に有効と判断した。

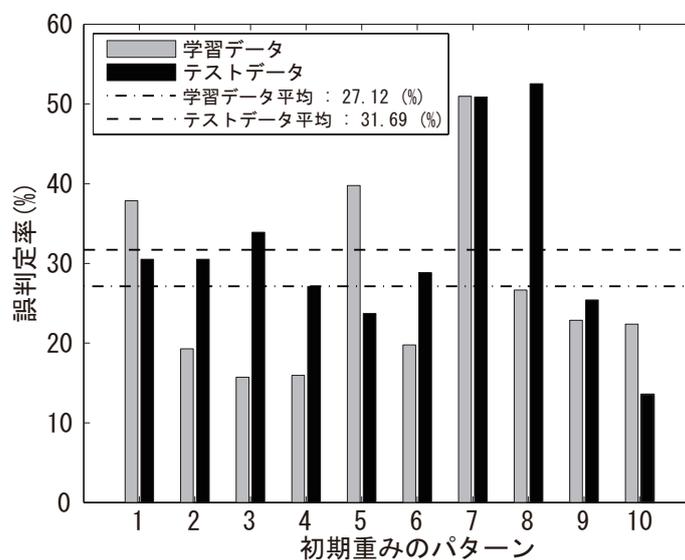


図69: 中間ニューロン数33個のネットワークにおける、重みの初期値が異なる場合での各学習結果と平均値

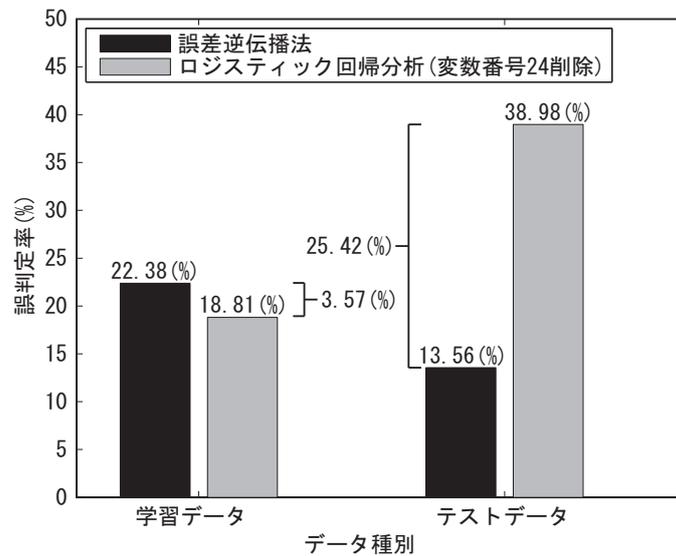


図 70: 分類結果

6.4.6 まとめ

Tweet データの分類に対して誤差逆伝播法とロジスティック回帰分析を比較したところ、テストデータの予測においてロジスティック回帰分析よりも誤判定率を 25.42(%) 低く予測することができた。これは約 15 個の判定の差であり無視できる差では無い。

くわえて精度以外にも、誤差逆伝播法が用意した全変数を解析できたのに対して、ロジスティック回帰分析では変数を削減しなければならなかった。削除した変数は公共交通機関の路線に関するものであり、今回の研究にとって重要な変数であるという仮説で作成した変数である。仮説に沿ったデータをそのまま分析できるという点でも誤差逆伝播法は有効な分類手法であると言える。

今回は 600 個のデータに対して分類実験を行ったが、今後はより大きいデータに対して誤差逆伝播法を適応し、最終的に作成を試みている情報提供サイトの構築を目指していく。また今回は分類精度にのみ着目し Tweet の分類に有効な手法について論じたが、今後は具体的な Tweet の分類結果も考察し、Tweet の分類に関する研究を更に進めていきたい。

6.5 重要判断基準抽出

6.5.1 解析目的

当解析は潜在学習法を使用して Tweet データから重要情報の検出を行うものである。

Tweet を自動で分類する研究は行われている [83]。しかし、どのような情報が重要なのかということ抽出することは行われていない。Twitter を分類し、また、分類に関する鍵となる情報を得ることは災害時の情報収集に有益であるため、当解析ではこれを目指す。

この目的を達成するため、解析は潜在学習法を用いて行った。この手法はニューロンの潜在性に着目して学習を行うもので、高い分類精度が期待できる。また、潜在性を解釈することで、ネットワークが何に着目したのかを解釈することができる。この手法を利用してこれまでに重要 Tweet の抽出を試み [73], [84] 汎化能力が一般的な手法と比べて高いことを確認した。しかし、解析に使用した Tweet 数が 600 件であり、また、重要と非重要な比率が同じ構成のデータを使用していた。このような比率の Tweet データは実際には存在しない。実際のデータは重要情報が非常に少ないものである。そこでより実際の環境に近いデータを使用し、潜在学習法が Tweet を正しく分類でき、重要情報の抽出が可能かを検証した。これは潜在学習法が非常に少ない重要 Tweet から分類に必要な特徴を得なければならない状況である。

6.5.2 解析データ

当解析で使用した 10000 件のデータは 2011 年 3 月 11 日 14 時 46 分 18 秒から 23 時 59 分 59 秒の間に発信された東日本大震災の Tweet データである。このうち 5000 件を分類モデル作成に使用し、残りの 5000 件を汎化能力検証に使用した。また、モデル作成用データは 3 区間に分割された。区間と比率は 70%を学習用、15%を学習中の検証用、そして 15%をモデルのテスト用である。このモデル作成用データは 145 件の重要データと 4855 件の非重要データから構成されている。

データの変数は形態素解析結果から得られた 41 個である。形態素解析ソフトウェアは「JUMAN」を用いた [79]。変数のうち 40 個は出現単語に関するもので、残りの一つは文字数である⁸。出現単語に関する変数とは単語ジャンルの出現個数である。単語ジャンルは JUMAN の辞書に搭載されているもので、出現した単語が「人に関する」や「ビジネスに関する」といったように、どの分野に関する単語なのかを表すものである。当解析では 40 個の単語ジャンルを使用し、それぞれのジャンルの単語が何回出現しているかをカウントしたものが 40 個の変数である。潜在学習法の間接ニューロンは 90 個を用いた、これは MATLAB の SOM toolbox の small の設定によるものである。

6.5.3 結果評価尺度

当解析ではモデルの精度評価に「Accuracy Ratio, AR」 [85] を用いた。これは誤判定率といった尺度では重要・非重要な比率が極端に異なるため精度評価が困難であるからである。

AR 値は $-1 \leq AR \leq 1$ の範囲で値を出力するものである。数値が大きいほど精度が良いことを意味する。AR は N_{IM} を重要 Tweet の個数、 N_{UN} を非重要 Tweet の個数、 Z_i を i 番目の重要な度合い、 Z_j を j 番目の非重要な度合い (これら度合いは潜在学習法からの出力である) とすると、

$$AR = \frac{2}{N_{IM}N_{UN}} \sum_{i=1}^{N_{IM}} \sum_{j=1}^{N_{UN}} I(Z_i - Z_j) - 1, \quad (6-1)$$

より求められる。ここで $I(x)$ 、

$$I(x) = \begin{cases} 1, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (6-2)$$

で定義される関数である。この尺度は重要 Tweet に対して潜在学習法が正しく重要度合いを出力しているかどうかを示すものである。

⁸解析時には各変数値は 0 から 1 の範囲に標準化され使用された。これは x' を標準化後の値とすると、 $x' = (x - \min(x)) / (\max(x) - \min(x))$ で求める手法である。

6.5.4 結果と考察

当解析では結果の比較のため、誤差逆伝播法、潜在性を使用しない潜在学習法、そして潜在学習法を用いて解析を行った。

表 31 はモデル作成用データのテスト区間での汎化能力を示したものである。モデル化はモデル作成用データを 10 パターン使用して行ったため、10 個のモデルの平均値を表 31 で示している。10 パターンのデータとは、5000 件のデータを 3 区間に分割した比率はそのまま、使用するサンプルをランダムに変更したものである。これはモデル化の際に偶然良い結果が得られるデータパターンが使用されることを防止するためである。モデル化の結果、平均 AR 値が 0.6956 のモデルが最も精度が高かった。これは潜在学習法により得られたものである。この結果は標準偏差 (Std dev) も最も低いものであった。次に精度が良かったモデルは誤差逆伝播法で AR 値は 0.6418 であった。続いて潜在性を使用しない潜在学習法という結果であった。この結果は SOM による事前学習が却って学習の妨げになったことを示唆する。

モデル化が完了したので、続いて 5000 件のモデルにとって初めて遭遇するデータを使用して、モデルの汎化能力を検証する。表 32 に検証結果を示している。この表より、潜在学習法が平均値、標準偏差、最小値、最大値で最も良い結果であることがわかる。続いて精度が良いモデルは誤差逆伝播法で、そして潜在性を用いない潜在学習法という結果であった。潜在性を使用しない潜在学習法はどの解析でも最も結果が良くないものであった。また、予測精度に加え、重要 Tweet の検出個数を各手法で確認を行った。結果は図 71 に示す通りであった。

モデル作成と検証が終了したので、続いて分類の鍵となる情報について抽出を行う。これには潜在性を確認することから、潜在学習法の結果を用いて実施する。図 72 は 41 個の入力ニューロンの潜在性を示している。入力ニューロンは入力変数に対応するため、最も大きい潜在性を持つ入力ニューロンが最も重要な入力変数という意味である。図 72 から 41 番目の「Tweet 中の文字数」が重要 Tweet と非重要 Tweet の分類に重

表 31: モデル作成用データのテスト区間での汎化能力

AR	手法		
	BP	SBP	PL ($r = 0.5$)
平均	0.6418	0.6403	0.6956
標準偏差	0.1429	0.1075	0.0876
最大	0.8034	0.7906	0.8278
最小	0.2758	0.4748	0.5649

表 32: 汎化能力検証用データでの汎化能力

AR	手法		
	BP	SBP	PL ($r = 0.5$)
平均	0.7849	0.7775	0.7943
標準偏差	0.0333	0.0336	0.0242
最大	0.8268	0.8322	0.8384
最小	0.7179	0.7159	0.7536

要であることがわかった。潜在性はこの入力ニューロン以外にも高い値を持っているものがあり、文字数だけで分類判定している訳ではないが、全体としてみると文字数の多い Tweet が分類の鍵を握っていることを示している。これは、文字数が多い Tweet ほど具体的な情報を含んでいることを示唆している。

当解析を通して、潜在学習法は判定する情報が少ないデータからでも特徴を十分に抽出できることがわかった。これは潜在学習法が潜在性に着目して学習を行うからである。潜在学習法は SOM を使用して事前に知識を獲得するものであるため、この事前知識から有効に知識を得ていることが分かる。この結果は潜在性を使用しない潜在学習法では学習がうまくできなかつたことから明らかである。潜在性に着目することにより、使用する知識の取捨選択が行われるが、SOM での結果を活かしてターゲットを予測するためには効果的な知識の取捨選択が必要であることを意味している。この取捨選択こそが潜在学習法の目的とするところである。これらのことにより、潜在学習法が震災発生時に実際に発信される状況のデータ分類と要因抽出に有効であることが示された。

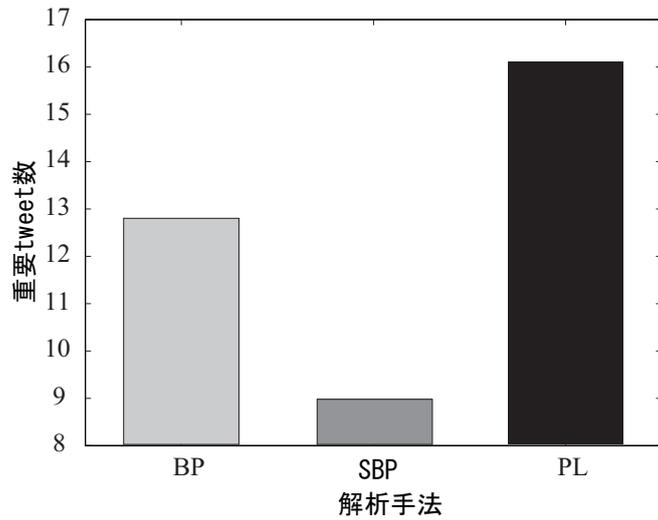


図 71: 重要 Tweet 検出数 (10 個モデルの平均値) (汎化能力検証用データ)

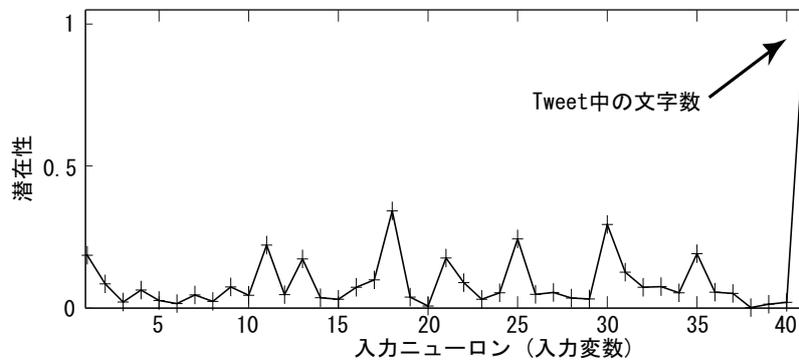


図 72: 入力ニューロン (入力変数) の潜在性 (モデル作成用データ) ($r = 0.5$)

6.5.5 まとめ

当解析では潜在学習法を用いて複雑なデータを解析することの効果について検証した。Twitter といった SNS は災害時に重要な役割を果たす。しかし、Tweet は冗長で重要情報が含まれた Tweet が少ないので効果的な活用が容易ではない。当解析以前に、潜在学習法は Tweet の分類に適用され分類精度が高いことを確認していたが、そのデータはサンプル数が少なく、そして重要・非重要な比率が偏っていないデータであった。そこで当解析ではサンプル数が大きく、重要・非重要な比率が偏ったデータを用いて潜在学習法の効果を確認した。解析の結果、汎化能力の高いモデルが作成できた。また、分類の中心的役割を果たす入力変数についても抽出できた。

しかし、10000 件というサンプル数は実際に災害発生時に発信される Tweet 数と比べると数が少なすぎる。今後の研究ではよりデータ数を増やしより現実的な条件で解析を行っていく。また、当解析では「文字数」が重要であることまでしか調査していないが、今後はより Tweet 内容の解釈を行い、災害時に人間がどのような情報を必要とするのかという人間行動に迫っていく。

7 企業行動原理抽出

本章では社是を解析対象として、社是に含まれている企業の行動原理抽出を試みたものについて述べる。社是は企業の最重要メッセージであり、全ての企業活動の原点はこの社是にあるといっても過言ではない。本章では企業の収益性に着目し、社是の違い（企業の行動原理の違い）について解析を実施する。解析は「東証二部上場企業の一部産業」、「東証一部上場企業の製造業」、「東証二部上場企業の製造業」を対象とした三つの解析を実施した。これら解析の結果概して「社会」という単語を含む社是を持つ企業が収益性が高いことが明らかになった。

7.1 解析概要

当解析は、企業の創業者あるいは企業そのものの行動原理・行動動機の抽出を試みたものである。ここでは東証二部上場の4業種を対象とする解析 [86]、東証一部上場の製造業企業を対象とする解析 [87]、そして東証二部上場の製造業企業を対象とする解析 [88] の三つの解析について述べる。

社是（経営理念や企業理念などとも称される [89]）は企業の意思を示す重要なメッセージであり、企業の存在理由・企業の行動指針などを企業外あるいは企業内に向けて発信している。企業最上位のメッセージであるため、ここで表明した事柄から逸脱した行動は企業として取りえない。経営理念の従業員への浸透と顧客・従業員満足の関係について研究した松葉によれば、経営理念の浸透度が顧客満足と従業員満足を両立させていると報告しており [90]、このように社是は企業活動に影響を及ぼす企業として重要な存在である。企業はこの社是に記された事柄を念頭に業務を行っていくため、企業の行動の違いは社是の内容の違いと捉えることができる。

本研究で社是を研究対象とするのは、各企業のパフォーマンスに違いが生じる理由について究極的に突き詰めると社是の違いがあるのではな

いかという仮説からである。当解析では企業の行動の違いと収益性の関係に着目し、社是の特徴について論じていく。

社是と業績の関係についてであるが、経営理念においては、2010年3月に経営哲学学会関東部会において「CSR、企業倫理、企業理念は本当に役に立つか？」と題されるシンポジウムが開かれ、企業パフォーマンスとの関係性を探った研究も報告された。従って、各企業にて企業理念や経営理念と同様に使用されている社是も企業パフォーマンスに影響を与えている可能性は非常に高く、また、分析対象を経営理念から広げることでより詳細な研究が行えると考えた。

これまで行われている社是に関する研究は、社是を内容や宣言対象別に類型化して行うもの [91], [92]、企業に対するアンケートより社是やその内容を収集し、業績との関係を研究したもの [93]、そして当解析でも実施するテキストマイニングを用いる研究である。

アンケートを用いた研究で飛田は、経営理念の言及先を「株主」、「従業員」、「株主と従業員の両方」、「株主と従業員のどちらにも言及していない」の四つに分類し、このうち株主について言及している経営理念を持つ企業の業績が、最も低いことを報告している [94]。テキストマイニングを用いた研究で小田らは、121社の製造業企業の社是を対象に形態素解析を実施し、得られた出現数10回以上の単語を非階層クラスタ分析を使用して解析を行い、社是を「従業員に対して心構え・あるべき姿を説いている経営理念を持つ企業のグループ」、「社会貢献・社会奉仕を訴える経営理念を持つ企業のグループ」、「世界への視点を持ち、活動を広げようとする経営理念を持つ企業のグループ」の三つのグループに分類し、これらのうち、「従業員に対して心構え・あるべき姿を説いている経営理念を持つ企業のグループ」が最も収益性が高いことを報告している。

7.2 社是のデータ化

社是はテキストデータであり、そのままでは計算機にて解析ができない。そこでテキストマイニングと呼ばれる手法を用いて数値データ化した。

テキストマイニングはテキストを対象に行う分析の総称である。テキストマイニングは、アンケートの自由記入欄に代表される、コメントデータの分析、ときには新聞社説 [95] など対象として活用されている。テキストマイニングを実施するためのソフトはさまざま存在しており、それぞれ特徴があるが、基本的には形態素解析と構文解析という技術を用いてテキストデータを解析する。

形態素解析は入力されたデータを「形態素」という言葉の最小単位に分解する技術、構文解析は文章の繋がりを明らかにする技術である。本研究では構文解析は行わず、形態素解析のみを「JUMAN」 [79] を用いて実行し、データを作成した。

本解析では 2013 年 10 月 31 日現在で東証一部、東証二部、マザーズに上場している企業を対象に社是を収集し、解析を行った。形態素・単語・単語カテゴリ・単語ドメインからなるデータを得た。単語カテゴリと単語ドメインは JUMAN の辞書にある概念で、単語の詳細情報を示している。例えば「時計」という単語は、カテゴリ：「人工物-その他」、ドメイン：「家庭・暮らし」となる。

解析対象となる社是は各企業の web サイトから収集を行った。このとき一定の条件に当てはまる場合には、社是そのものではなく別のメッセージを社是として収集した。一例ではあるが、社是が複数の文言から成り立っている場合は最初に述べられている文言を、社是が抽象的で且つその社是の説明文が存在する場合はその説明文を、社是よりも上位のメッセージが存在する場合はその上位のメッセージを、社是が web サイトに記載されていない場合で経営理念などが存在する場合は経営理念を、それぞれ社是として収集した。

以下、東証二部上場の 4 業種に関する解析は「業績一覧：国内会議 [8] (参考文献 [86])」記載の論文を、東証一部上場の製造業企業に関する解析は「業績一覧：国内会議 [5] (参考文献 [87])」記載の論文を、東証二部上場の製造業企業に関する解析は「業績一覧：国内会議 [1] (参考文献 [88])」記載の論文を転載 (構成の入れ替えや一部加筆・修正を行っている) して述べる (転載について当該学会より許諾を得ている)。

7.3 東証二部上場の4業種を対象とする解析

7.3.1 解析目的

当解析は、社是が企業の利益（企業パフォーマンス）に与える影響を研究したものである。社是を唯一の入力としてクラスタリングすると共に、そのクラスタリング結果に至った原因から、どのような社是が企業パフォーマンスに影響を与えるのかを探った。解析対象は東京証券取引所市場第二部（東証二部）に上場している企業のうち、4業種を分析対象とした。

当解析では収集した社是をテキストマイニングにて処理した後、クラスタリングとその理由の解明（入力変数からの特徴抽出）を行った。テキストマイニング手法を用いた経営理念と企業パフォーマンスに関する先行する研究として2010年の小田らの研究があげられる [93]。小田らは製造業121社の経営理念を分析対象としてテキストマイニングを行い、頻出単語を抽出、その単語情報を入力として非階層クラスタ分析を行い、企業を「従業員に対して心構え・あるべき姿を説いている経営理念を持つ企業のグループ」（グループ1）、「社会貢献・社会奉仕を訴える経営理念を持つ企業のグループ」（グループ2）、「世界への視点を持ち、活動を広げようとする経営理念を持つ企業のグループ」（グループ3）、の3グループに分類し、そしてこの三つのグループ中、グループ1が最もパフォーマンスが良く、グループ2が最も悪かったと報告している。ただし、この研究ではグループ数が三つとなったが、これは小田らの判断によるものである。

そこで本研究では企業の分類に分析者の恣意が入ることを排除する目的で、自動で分類を行った。この分類は機械的に限界まで分類したのではなく、「反復情報量最大化法, Gradual Information Maximization, GIM」を用いて知能処理的に行った。

7.3.2 分析

当解析では2013年10月31日現在で東証二部に上場している企業の内、「機械」、「金属製品」、「建設業」、「不動産業」の4業種を対象に分析を行った。この時点で、これら4業種の東証二部に上場している企業数は120社あるが、社是が公開されていないなどの理由より、本研究の対象企業は全部で90社であった。本分析の手順はデータ作成も含めて以下の通りである。

1. データ作成

- (a) 各企業のホームページより社是の収集を行う。ただし、「社是」という名称ではない、「社是よりも上位の企業発信メッセージがある」といった場合は、そのメッセージ(経営理念や創業の精神など)をデータとして用いた。また、メッセージが箇条書きなどにより複数個から構成されている場合は、最初の1メッセージのみを分析対象とした。
- (b) テキストマイニングを行い、メッセージで使用されている単語の品詞、カテゴリ名、ドメイン名の出現頻度を抽出する⁹。カテゴリ名、ドメイン名とはテキストマイニングソフトウェア「JUMAN」の出力から得られるものであり、例えば「学校」という言葉は、カテゴリ「場所-施設」、ドメイン「教育・学習」という具合に結果が出力され、分析対象の言葉の詳細な説明となっている。

2. クラスタリングを行う。

- (a) 社是データをGIMに入力し学習させ、社是(企業)を分類する。
- (b) 情報量を基準に、クラスタリングの原因となった事柄(重要入力変数)を見つけ出す。

3. クラスタリング結果と利益との関係を調べる。

⁹解析時には各変数値は0から1の範囲に標準化され使用された。これは x' を標準化後の値とすると、 $x' = (x - \min(x)) / (\max(x) - \min(x))$ で求める手法である。

- (a) 発見した重要変数を手がかりに、どのような言葉が各グループに含まれているのかを調べる。
- (b) 得られたグループ毎に、そのグループに属する企業の経常利益の平均値を計算する。なお、経常利益は金融庁の電子開示システム (EDINET) より収集した。
- (c) 各グループの平均経常利益を基準として、各グループの順位付けを行いその業種について傾向を考察する。

それでは分析対象とした各業種について、それぞれ個別に分析結果を見ていくことにする。なお、分析結果の図「Labels」中に記載のある数字は、それぞれ企業を表している。GIMを用いて分析を行うと、大きく分けて「U-Matrix」と「Labels」という二つの図を出力として得ることができる。どちらも出力層のニューロンを表した図であり、大抵の場合ニューロンは六角形の図形で表される。「Labels」はそのまま出力層と対応する、一方「U-Matrix」は隣り合う出力層のニューロンとの距離を加えて表したものである。「Labels」にはデータの観測値の名称が記載されており、この図からどのデータがどこに分類されたかが確認できる。「U-Matrix」は主に色で表現され、暖色が「隣のニューロンと大きく離れていること」を、寒色が「近いこと」を表しており、暖色のニューロンを境界線と見て各企業がそれぞれグループ分けされていると見ることができる。つまり、「U-Matrix」でクラスタリングの状態を、「Labels」で具体的にどのデータかということを見ていくことになる。

7.3.3 結果と考察

7.3.3.1 業種：機械

業種「機械」の分析対象企業数は39社であった。図73はGIMによる分類結果を示している。図73(a)の結果より、この業種の企業は三つのグループに分割されたことが分かる。分類結果を具体的に表したものが図73(b)である。図74はこの分類結果に重要な役割を果たした入力ニューロンの発火確率を表している。

図 74 によると、分割の原因となった変数は変数番号 31 のカテゴリ名「人工物-その他」であった。グループ 1 は重要変数のカテゴリに属する単語が出現したグループ、グループ 2 も重要変数のカテゴリに属する単語が出現したグループ、そしてグループ 3 は出現が無かったグループである。これらのうち、グループ 2 が最もパフォーマンスが高く、グループ 3、グループ 1 と続いた。具体的な単語であるが、グループ 2 は「機器」「機械」「薄膜」「製品」、グループ 3 は「物」「製品」であった。なお、この業種の分析に使用した β 値は 10 である。

7.3.3.2 業種：金属製品

業種「金属製品」の分析対象企業数は 18 社であった。図 75 は GIM による分類結果を示している。

この業種は図 75 (a) より、二つのグループに分割されたことが分かる。入力ニューロンの発火確率は図 76 に示す通りである。

図 76 によると、分割の原因となった変数は変数番号 56 のドメイン名「教育・学習」であり、該当する単語は「社会」であった。グループ 1 は重要変数のドメインに属する単語が出現しなかったグループ、グループ 2 は出現したグループで、パフォーマンスはグループ 2、グループ 1 の順であった。この業種の分析に使用した β 値は 20 である。

7.3.3.3 業種：建設業

業種「建設業」の分析対象企業数は 22 社であった。GIM による分類結果 (図 77) によると、この業種の企業は四つのグループに分割されたことがわかる。入力ニューロンの発火確率 (図 78) を見ると、変数番号 56 のドメイン名「教育・学習」を理由として分割されたことが分かる。

グループ 1 はドメイン名「教育・学習」に属する単語の出現の無いグループ、グループ 2 は出現のあるグループ、グループ 3 は出現のあるグループ、グループ 4 は出現の無いグループであった。出現した単語は、グループ 2 は「社会」、「福祉」、「通信」、グループ 3 は「社会」、「通信」であり、パフォーマンスは高いグループから、グループ 3、グループ 2、グ

表 33: 各グループとパフォーマンスの関係

パフォーマンス	業種			
	機械	金属製品	建設業	不動産業
高	G2 (あり)	G2 (あり)	G3 (あり)	G1 (あり)
			G2 (あり)	
	G3 (あり)		G4 (なし)	
低	G1 (なし)	G1 (なし)	G1 (なし)	G2 (なし)

グループ4、グループ1と続いた。この業種の分析に使用した β 値は10である。

7.3.3.4 業種：不動産業

業種「不動産業」の分析対象企業数は11社であった。図79に示すように、GIMによると不動産業のデータは二つのグループに分割された。その原因となった変数は変数番号56のドメイン名「教育・学習」であり(図80)、「社会」という単語がそれに当たる。グループ1は単語出現のあったグループ、グループ2は出現のなかったグループで、パフォーマンスは高い方から、グループ1、グループ2と続いた。この業種の分析に使用した β 値は10である。

7.3.3.5 各グループと収益

ここまでの分析で各企業の社是はその違いにより複数のグループに分類されることが分かったが、これらのグループと収益との関係を照らし合わせると表33のようになる。「あり」、「なし」とあるのは、重要変数に属する単語の出現があったか・なかったかを示している。表33から、重要変数を含む単語が社是に使用されている企業が、高いパフォーマンスを出していることが分かる。表33の関係を具体的に経常利益の値(単位：千円)で表したものが、表34である。

表 34: 各グループと平均利益 (千円) の関係

パフォーマンス	業種			
	機械	金属製品	建設業	不動産業
高	1 292 211	1 232 800	4 055 800	238 070
			836 900	
	1 090 700		496 437	
低	671 830	795 480	297 110	203 490

表 35: 重要変数出現グループ内での共通単語

業種				
機械	金属製品	建設業	不動産業	
技術 5	社会 5	社会 8	社会 4	
貢献 5	貢献 5	貢献 8	企業 2	
社会 4	環境 3	技術 6	住 2	
製品 4	物 2	企業 4	提供 2	
当社 4	発展 2		空間 2	
提供 3	豊かだ 2		貢献 2	

7.3.4 まとめ

ここまで4業種を個別に分析してきたが、今回分析対象とした四つの業種に共通して重要変数を含むグループの方が含まないグループよりも企業パフォーマンスが高いという結果を得た。具体的な出現単語であるが、4業種中3業種において共通して「社会」という単語の出現が見られ、また、残り1業種(「機械」)の出現単語は「機器」や「機械」など、その業種に直接的に関わる単語であった。

続いて重要変数単語が出現するグループ内で、どのような単語が共通して出現しているのかを名詞と形容詞に絞って見てみることにする(表35)。この表で各単語の右に記載のある数字であるが、これはその単語が何社重複して出現した数かということを表している。

また、重要変数含有単語を含む重複数上位3件までの単語を表示している。なお、いくつかの業種で「重要変数出現あり」のグループが複数個出現したが、この表では「出現あり」グループとしてまとめて結果を掲載している。

表 35 より各業種共通の単語として「社会」と「貢献」という単語が見られた。このことより、企業パフォーマンスの良い企業の社是は「社会貢献」を訴える社是であることが読み取れる。このタイプの社是は鳥羽らの研究 [91] によると「方針型」という類型に属し、対外的効果を狙うと同時に対内的効果も有するという。また、コーポレート・アイデンティティといった考え方もこのタイプであり、更にアメリカの経営理念はこの「方針型」に最も近いと報告している。海外発祥の考え方がよく取り入れられる近年の状況から鑑みても、今回は4業種のみ分析ではあるが「東証二部を構成する企業のうち収益の良い企業は、その社是で「社会貢献」を訴えている」という結果は妥当性がある。

本研究の結果をまとめると以下の通りである。

- 重要変数を含む単語を社是で使用している企業は、使用していない企業と比較してパフォーマンスが良かった。
- パフォーマンスの良い企業の社是は、各社共通して使用される傾向の高い単語から判断して「社会貢献」を訴える社是であることが分かった。

今回の研究と小田らの研究は分析対象業種が異なるので完全に比較することは難しいが、小田らの「社会貢献・社会奉仕を訴える経営理念を持つグループが最もパフォーマンスの低いグループであった」という報告と異なった結果であり、興味深い結果が得られた。残念ながら今回得られた結果から小田らの研究結果との違いについて考察することは、得られたグループが「社会貢献を訴えるグループ」だけということがあり比較するための結果に乏しく踏み込めなかった。

今後はまだ分析できていない東証二部の他の業種を分析し、「社会貢献」以外のグループが得られるのかを調査する。そして先行研究との比較を考察していきたい。また、東証マザーズそして東証一部と分析範囲を広げ、最終的に日本の全上場企業の傾向について論じていきたい。

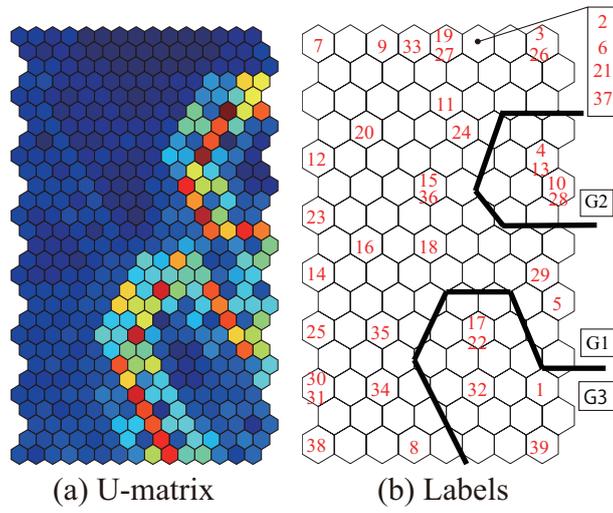


図 73: 業種：機械、 U-Matrix と Label

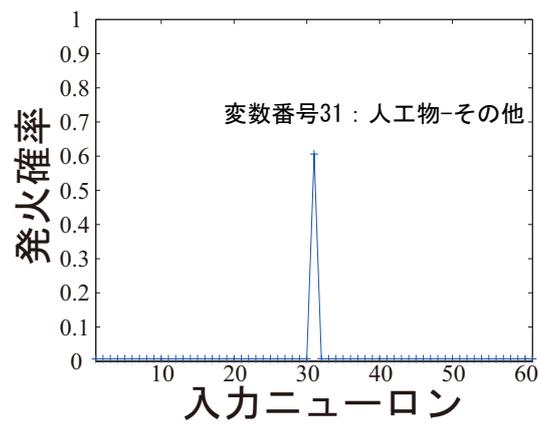


図 74: 業種：機械、 重要変数

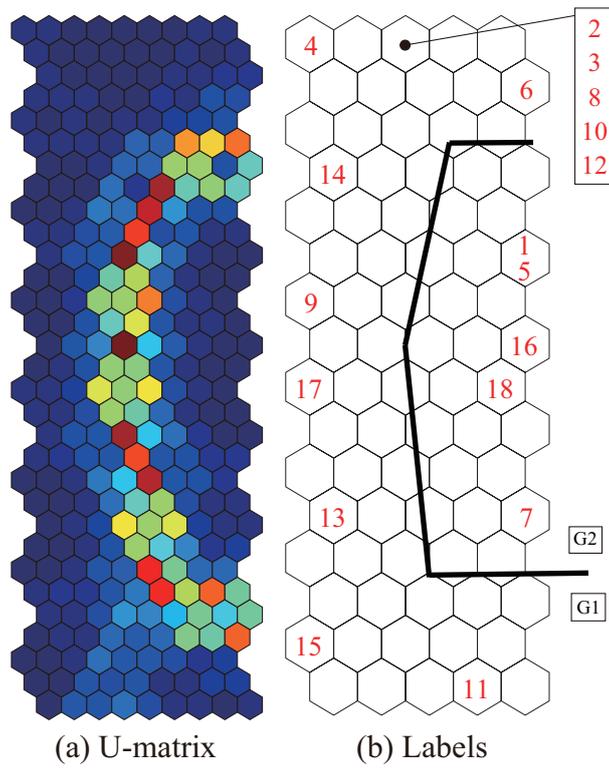


図 75: 業種：金属製品、U-Matrix と Label

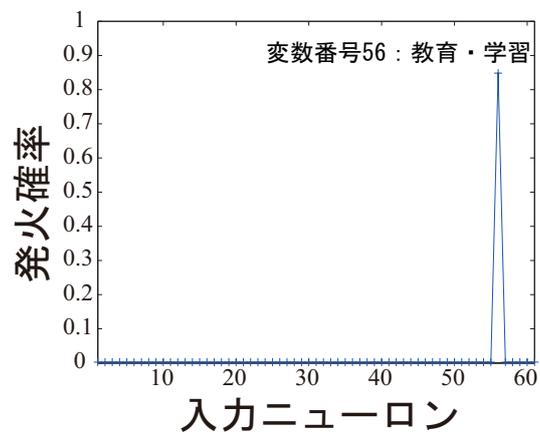


図 76: 業種：金属製品、重要変数

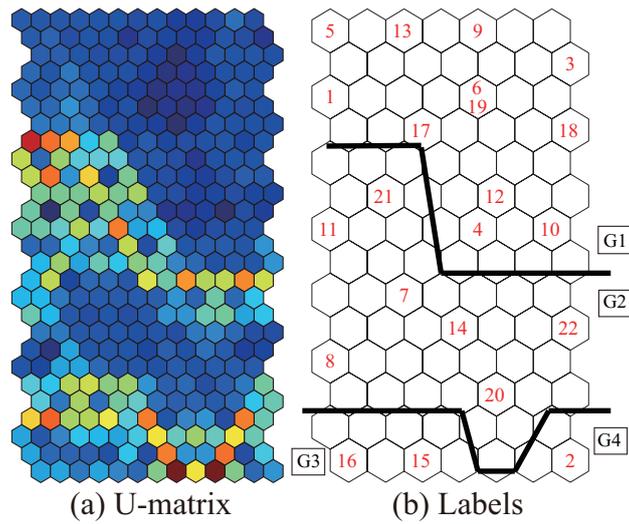


図 77: 業種 : 建設業、U-Matrix と Label

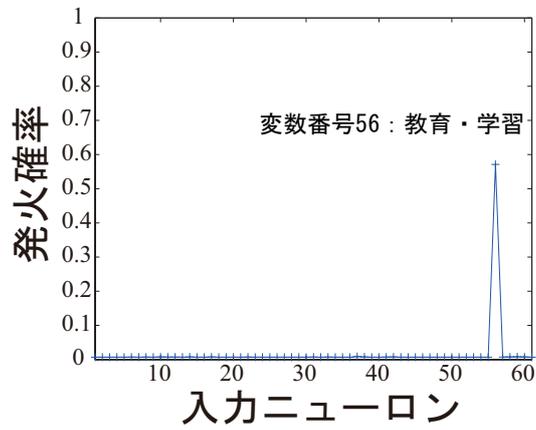


図 78: 業種 : 建設業、重要変数

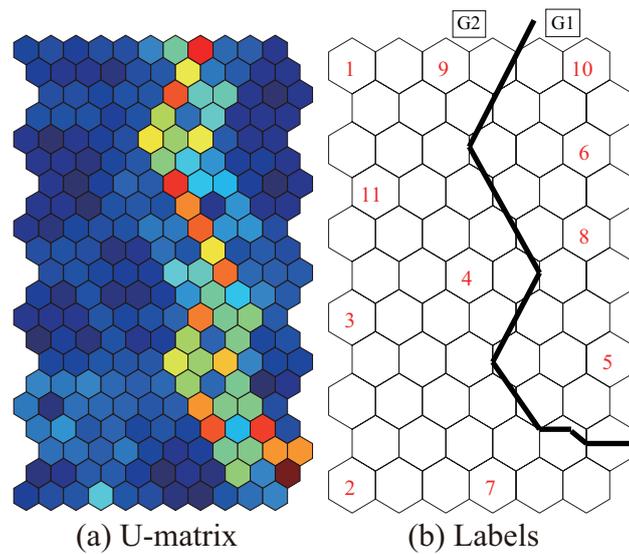


図 79: 業種 : 不動産業、U-Matrix と Label

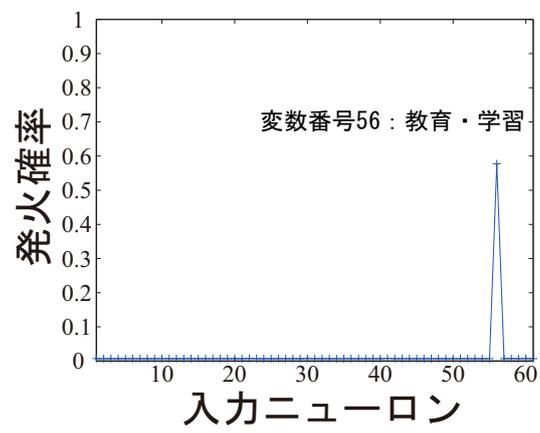


図 80: 業種 : 不動産業、重要変数

7.4 東証一部製造業企業を対象とする解析

7.4.1 解析目的

当解析は東京証券取引所、市場第一部（東証一部）に上場している製造業企業の社是を対象に解析を行ったものである。また、前回の解析と異なり、内容語のみを対象として解析データを作成した。

7.4.2 データ

7.4.2.1 解析対象社是の収集

当解析では 2013 年 10 月 31 日現在で東京証券取引所、市場第一部（東証一部）に上場している製造業の企業、711 社の社是を対象に解析を行った。製造業は「証券コード協議会」の定めた「業種別分類項目及び業種コード」 [96] の大分類に記載されている業種で、表 36 の中分類に記載のある業種から構成されている。

表 36: 業種別分類項目及び業種コードの定める製造業
(証券コード協議会の web サイト [96] より一部引用)

業種	
大分類	中分類
	食料品
	繊維製品
	パルプ・紙
	化学
	医薬品
	石油・石炭製品
	ゴム製品
製造業	ガラス・土石製品
	鉄鋼
	非鉄金属
	金属製品
	機械
	電気機器
	輸送用機器
	精密機器
	その他製品

表 37: 作成した変数一覧

番号	変数名	番号	変数名	番号	変数名	番号	変数名
1	名詞	11	人工物-衣類	21	抽象物	31	地名:国
2	形容詞	12	人工物-乗り物	22	形・模様	32	文化・芸術
3	動詞	13	人工物-金銭	23	色	33	レクリエーション
4	人	14	人工物-その他	24	数量	34	スポーツ
5	組織・団体	15	自然物	25	時間	35	健康・医学
6	動物	16	場所-施設	26	企業名	36	家庭・暮らし
7	植物	17	場所-施設部位	27	企業グループ名	37	料理・食事
8	動物-部位	18	場所-自然	28	造語	38	交通
9	植物-部位	19	場所-機能	29	四字熟語	39	教育・学習
10	人工物-食べ物	20	場所-その他	30	地名:日本	40	科学・技術

7.4.2.2 解析用データの前加工

当解析では表 37 に示す解析用データを作成した。しかしデータの変数を確認したところ、出現頻度が少ない変数が多く存在していることが分かった。そこで出現頻度が 10 回より少ない 16 個の変数を削除し、合計 28 個の変数を解析に使用することにした¹⁰ (解析に使用した変数一覧を表 38 に示す)。

表 38: 解析に使用した変数一覧

番号	変数名	番号	変数名	番号	変数名
1	名詞	11	場所-機能	21	健康・医学
2	形容詞	12	場所-その他	22	家庭・暮らし
3	動詞	13	抽象物	23	料理・食事
4	人	14	形・模様	24	交通
5	組織・団体	15	数量	25	教育・学習
6	動物-部位	16	時間	26	科学・技術
7	人工物-食べ物	17	企業名	27	ビジネス
8	人工物-その他	18	企業グループ名	28	政治
9	自然物	19	文化・芸術		
10	場所-施設	20	スポーツ		

¹⁰解析時には各変数値は 0 から 1 の範囲となるように標準シグモイド関数を用いて標準化され使用された。これは x' を標準化後の値とすると、 $x' = 1/(1 + e^{-\hat{x}})$ より求められる。ここで、 $\hat{x} = (x - \bar{x})/\sigma_x$ であり、 \bar{x} は x の平均値、 σ_x は x の標準偏差である。

7.4.2.3 収益性データ

本研究では社是データに加え、企業の収益性に関するデータ「総資産利益率, ROA」も使用する。本研究で使用した ROA は、

$$\text{ROA} = \frac{\text{経常利益}}{\text{総資産}}, \quad (7-1)$$

より算出した。なお、使用した経常利益と総資産は金融庁の電子開示システム (EDINET) を用いて収集した。

7.4.3 解析手法

当解析では「Selective Potential Learning, SPL」 [97] を使用して解析を行う。SPL は潜在学習法の一つであり、構造は潜在学習法と同様に 2 段階の学習段階を有している (図 81 にネットワーク概要図を示す)。主な違いは潜在性の得方である。SPL でも潜在学習法と同様に第 1 段階で計算される潜在性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が求め方が異なっている。 $w_{j,k}$ 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、

$$V_k = \exp \left(R \sum_{j=1}^J (w_{j,k} - w_k)^2 \right) \quad (7-2)$$

で定義する。ここで R は係数であり、潜在性を強調する役割を持っている。本研究では $R = r\sqrt{J}$ とした (r については後記する)。また、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

$$w_k = \frac{1}{J} \sum_{j=1}^J w_{j,k} \quad (7-3)$$

から計算される。 k 番目の入力ニューロンの潜在性 $p(k)$ は、

$$p(k) = \frac{V_k}{\sum_{k=1}^K V_k} \quad (7-4)$$

より求められる。ここで K は入力ニューロンの個数である。

このようにして求められた潜在性は自己組織化マップ (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、

$$\|\mathbf{x}^s - \mathbf{w}_j\| = \sum_{k=1}^K (x_k^s - w_{j,k})^2 \quad (7-5)$$

により入力データと重みの距離 (差の自乗和) が求められるが、SPL では、

$$\|\mathbf{x}^s - \mathbf{w}_j\| = \sum_{k=1}^K p(k; r) (x_k^s - w_{j,k})^2 \quad (7-6)$$

により距離を求める。このとき $p(k; r)$ は、パラメータ r の時の $p(k)$ である。 r は第1段階での学習の繰り返し回数を意味しており、 r が増加することにより潜在性も増加する。以上が SPL の概要である。

SPL は潜在性の高い入力ニューロンに着目して学習を行うニューラルネットワークである。さまざまな状況でも発火するニューロンとはすなわち、重要な役割を担っているニューロンであるという意味であり、これを調べることで重要変数について解釈ができる。

7.4.4 結果と考察

7.4.4.1 SPL による社是の分類の結果

解析用データを SPL にて解析した結果を図 82 (b) に示す。この図はパラメータ r が 16 の際の SPL による社是の分類結果を示している ($r = 16$ を採用した理由は 7.4.4.3 で述べる)。

ここで示している図は 2 次元の図であるが、色の違いによって奥行が表現され 3 次元の図となっている。色は奥行について「高さ・低さ」を意味しており、寒色は高度が低く、暖色は高度が高いことを示している。つまり暖色で示された箇所がそうでない箇所と比べて断絶が大きい箇所を示しており、これはすなわち境界線を意味する。

図 82 (b) では 1 本の境界線が観測されるため、社是は二つのグループに分かれていると言える。今回は境界線より上部を「グループ 1」、下部を「グループ 2」と命名した。今回用いた社是の業種がどちらのグループに何社分けられたのかを、表 39 に割合にて示す。灰色で網掛けにしている箇所は一方のグループと比較して所属業種の割合が大きい箇所である。

ここで比較のため、自己組織化マップによる分類結果を図 82 (a) に示す。図 82 (a) 中には境界線が見られるものの、SPL と比較して境界線が明確ではなく、まだ多くの特徴抽出が行える余地がある印象を持つ結果である。この比較より、社是の解析には SPL が有効であることが言える。

7.4.4.2 SPL による分類結果の要因

図 83 は図 82 (b) の分類が得られた際の潜在性を示したもので、この分類結果に大きな役割を果たした変数 (要因) を意味している。この図より、最も潜在性が高く、本分析結果で最も重要な変数は 25 番変数 (教育・学習) であることが分かる。

7.4.4.3 SPL によるターゲットの予測結果

ここまで 25 番変数に基づいて社是が分類された結果を示してきたが、ここではその分類に用いた知識を使用したターゲットの予測結果を示す。本研究でのターゲットは収益性の高低であり、ROA が平均値 (約 4.9663) 以上であれば「高い」、平均値よりも低ければ「低い」とした。

図 84 はターゲットの予測結果を示したものである。ターゲットの予測は、「BP」、「SPL」、「SOM で学習した知識を BP で用いる方法, SBP (潜在性を使用しない SPL と同義)」の三つの方法により実施し、予測誤差を比較した。予測は 10 回繰り返して行い、ここではその平均値を示している。予測誤差は、SPL による予測が最も小さく、 $r = 16$ の時に最小誤差である 0.4327 を得た。

このように $r = 16$ の際に最小誤差を得られたことが、図 82 (b) で $r = 16$ の際の図を示し、それを SPL による社是の分類結果とした理由である。

7.4.4.4 分類結果の構造

以降の節では、SPLにより得られた結果を用いて、東証一部上場の製造業企業の社是について解釈を行う。解釈にあたって製造業企業に対する分類を一つ導入する。表 40 は 2011 年 3 月に経済産業省が公表した「我が国の工業」 [98] に記述されている「産業 3 類型」を示している。これは製造業を構成する企業を三つに分類したものである。今回、表 39 の SPL による分類結果を解釈するにあたり、この産業 3 類型を用いる。「我が国の工業」は日本標準産業分類の大分類の製造業に属する企業を対象に行われる工業統計調査について記述されたものである。この日本標準産業分類の大分類は、製造業の社是を収集する際に用いた、業種別分類項目及び業種コード (表 36) の大分類と大分類名こそ同一の「製造業」であるが、中分類は同一ではない。しかし、業種別分類項目及び業種コードも日本標準産業分類を基に分類を定めているため、本研究では製造業の社是の解釈に産業 3 類型を用いることにする。

表 39 中の列「産業 3 類型」は社是データに含まれる中分類業種と産業 3 類型 (表 40) を対応させたものである。産業 3 類型は今回解析対象となった製造業の中に、生活関連型産業は 3 業種、加工組立型産業は 4 業種、基礎素材型産業は 9 業種存在している。このことより表 39 を判断すると、グループ 1 には生活関連型産業が 100%、加工組立型産業が 75%、基礎素材型産業が約 56%存在しており、グループ 2 には加工組立型産業が 25%、基礎素材型産業が約 33%存在している事が分かる。つまりグループ 1 は産業 3 類型を構成する産業が、多数属するグループ、グループ 2 は少数属するグループであった。なお、石油・石炭製品企業の社是は各グループ中での比率が同じため、基礎素材型産業の割合の計算には含めておらず、合計が 100%にはならない。グループ間の差が、属する産業による違いではなかったという事より、製造業企業の社是は、産業の違いとは別の何らかの事柄 (事柄 X と命名する) により特徴づけられていることが示された。

7.4.4.5 分類結果と重要変数との関係

ここではSPLにより最も重要とされた25番変数(教育・学習)に着目し、前節で発見された事柄Xも含め、分類結果の解釈を行う。この変数に関する単語の出現について各グループで調査したところ、グループ1に所属する企業の社是には25番変数に関する単語は出現せず、グループ2に所属する企業の社是では100%で出現していた。グループ2の社是に含まれる25番変数に関する単語は“化学”、“学ぶ”、“社会”、“社会資本”、“福祉”、“自転”、“解析”、“追究”、“通信”であった。これらの単語のなかで“社会”は333回出現しており最多の出現回数であった。これにより25番変数の内容は単語“社会”であると判断し、グループ1とグループ2に分類される社是の違い、つまり事柄Xは、社是中の“社会”という単語の含有有無であると判断出来る。

7.4.4.6 分類結果と収益性との関係

続いてこの分類結果と各グループに属する企業の収益性(ROA)を比較してみた。各グループのROAを計算したところ、グループ1所属企業の平均ROAは約5.2718、グループ2では約4.5526という結果を得た。なおこの結果は、両側検定のt検定を行ったところ、 $t(709) = 2.3751$ 、 $p < .05$ で有意であった。

表 39: 各グループに所属する業種の割合

業種名	データ中の企業数	グループ1	グループ2	産業3類型
食料品	62	0.6935	0.3065	生活関連型産業
繊維製品	36	0.6389	0.3611	生活関連型産業
パルプ・紙	10	0.200	0.8000	基礎素材型産業
化学	103	0.5243	0.4757	基礎素材型産業
医薬品	34	0.7941	0.2059	基礎素材型産業
石油・石炭製品	10	0.5000	0.5000	基礎素材型産業
ゴム製品	9	0.7778	0.2222	基礎素材型産業
ガラス・土石製品	28	0.5714	0.4286	基礎素材型産業
鉄鋼	24	0.2083	0.7917	基礎素材型産業
非鉄金属	18	0.5556	0.4444	基礎素材型産業
金属製品	31	0.4839	0.5161	基礎素材型産業
機械	92	0.5870	0.4130	加工組立型産業
電気機器	135	0.5407	0.4593	加工組立型産業
輸送用機器	58	0.6207	0.3793	加工組立型産業
精密機器	22	0.4545	0.5455	加工組立型産業
その他製品	39	0.7436	0.2564	生活関連型産業

表 40: 産業3類型(我が国の工業 [98] より引用し、一部著者改変)

業種	産業3類型
食料・飲料	生活関連型産業
繊維・衣類	生活関連型産業
家具・装備品	生活関連型産業
印刷・同関連業	生活関連型産業
なめし革・同製品	生活関連型産業
その他	生活関連型産業
一般機械	加工組立型産業
電気機械	加工組立型産業
輸送機械	加工組立型産業
木材・木製品	基礎素材型産業
パルプ・紙・紙加工品	基礎素材型産業
化学工業	基礎素材型産業
石油・石炭製品	基礎素材型産業
プラスチック製品	基礎素材型産業
ゴム製品	基礎素材型産業
窯業・土石製品	基礎素材型産業
鉄鋼業	基礎素材型産業
非鉄金属	基礎素材型産業
金属製品	基礎素材型産業

7.4.5 まとめ

本研究では東証一部上場の製造業企業の社是を対象に、新しいニューラルネットワークである SPL を用いて解析を行った。解析の結果、社是は社会という単語を含まないグループ (グループ1) と、含むグループ (グループ2) に分割されること、そしてグループ1の方が多数派であることが分かった。またこの分類結果を収益性と照らし合わせたところ、グループ1の企業の方がグループ2の企業よりも収益性が高いことが分かった。

本研究では2013年10月31日現在で東証一部に上場されている製造業の企業のみを対象として解析を行った。今後は東証マザーズなど、他上場部の製造業や、他業種の社是を解析するなどし、社是の解析を更に進めて行く。

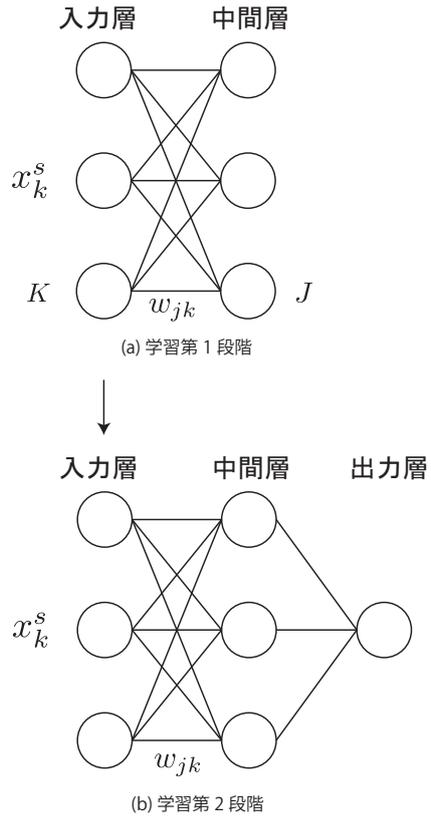


図 81: SPL 概要図

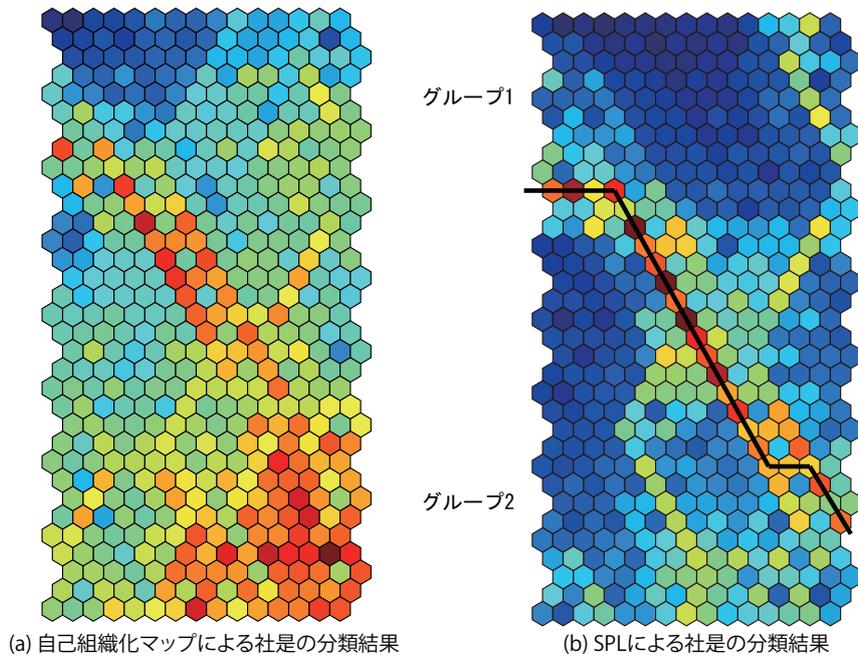


図 82: 社是の分類結果

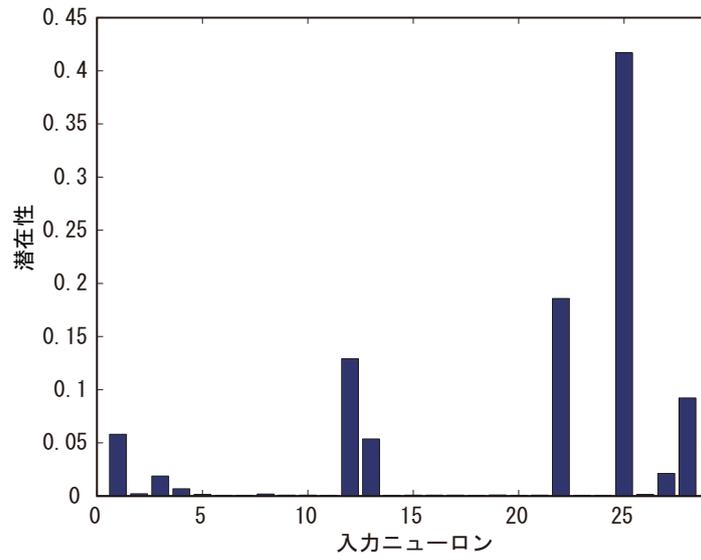


図 83: SPL により得られた重要変数についての出力

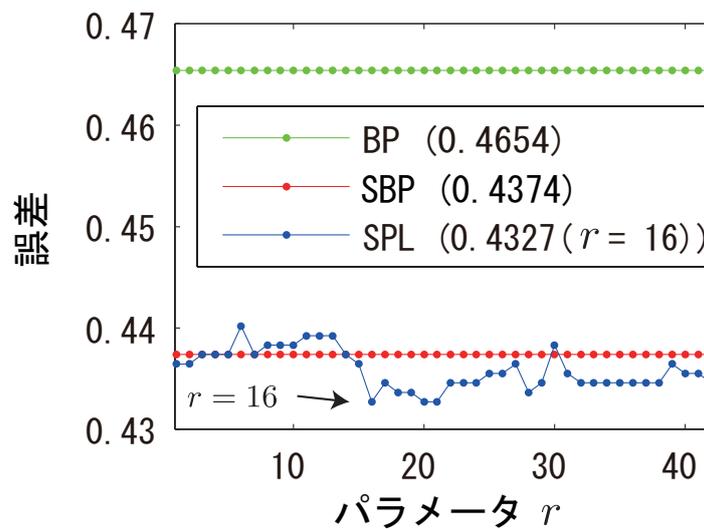


図 84: ターゲットの予測誤差

7.5 東証二部製造業企業を対象とする解析

7.5.1 解析目的

本論文は社是と企業パフォーマンスの関係について TF-IDF と共起語を軸として解析し論じたものである。

ここまで二つの社是データの解析事例を示してきた。そして東京証券取引所、市場第二部 (東証二部) 上場企業を対象とした研究 [99], [86], [100] では一貫して、「社会」という単語が企業パフォーマンスに影響を与えていることを報告した。

しかし、これらの研究にはデータと解釈に改良の余地が残されていた。それは形態素解析の結果から各単語のジャンルを抽出し、その出現頻度 (Term Frequency, TF) を用いて解析を実施していたことと、解釈に共起関係を用いていないことである。TF のみを対象としている従来の研究では、単語の各社是中での出現状態を考慮していないため、分析や解釈が単純に出現頻度が多いものに強く影響を受けている可能性を払拭することができない。

そこで、本研究では出現頻度を出現状態で重み付けする「文書頻度逆数 (Inverse Document Frequency, IDF)」を用いてデータ上の問題点を改良し、また、解釈に共起関係を加えて社是と企業パフォーマンスの関係を新たに解析した。

7.5.2 解析対象企業と解析データ

解析対象とした社是は東京証券取引所、市場第二部 (東証二部) 上場の製造業企業より収集を行った 214 社の社是である。

JUMAN から得られたデータは続いて内容語 (名詞・形容詞・動詞) を対象として品詞・単語カテゴリ・単語ドメインの三つの変数ジャンルについて単語の集計が行われた。各変数ジャンルは品詞が 3 個、単語カテゴリが 28 個、単語ドメインが 13 個から構成されている。集計結果には単語出現数 (TF) を出現状態 (IDF) を加味して重み付けした TF-IDF を用いた。また、この TF-IDF に対し「コサイン正規化」 [101] を用いて正

表 41: 解析データの変数一覧

変数番号	変数ジャンル	変数名	変数番号	変数ジャンル	変数名
1		名詞	20		文化・芸術
2	品詞	形容詞	21		スポーツ
3		動詞	22		健康・医学
4		人	23		家庭・暮らし
5		組織・団体	24		料理・食事
6		動物-部位	25	単語ドメイン	交通
7		人工物-衣類	26		教育・学習
8		人工物-金銭	27		科学・技術
9		人工物-その他	28		ビジネス
10		自然物	29		メディア
11	単語カテゴリ	場所-施設	30		政治
12		場所-機能			
13		場所-その他			
14		抽象物			
15		数量			
16		時間			
17		企業名			
18		企業グループ名			
19		四字熟語			

規化を行った。これにより、各変数ジャンルの変数値の2乗和が1となるように正規化された。本研究では変数ジャンルが三つ存在するため、全変数値の2乗和は3となる。

以上の手順により社是は数値化され、44変数からなる解析対象データを得たわけであるが、本研究ではこの214社からなる社是データをさらに絞り込み解析を実施した。

絞り込みの条件は「企業パフォーマンスが全東証二部製造業企業平均の標準偏差一つ分以上または以下である企業」かつ「全変数の値が0ではない」かつ「全企業で使用されている変数」である。三つの絞り込み条件の前半二つが企業の絞り込み、そして後半が使用変数の絞り込みである。「全変数の値が0の企業ではない」という条件は社是が外国語で構成されている場合に起こるものである。これらの条件により絞り込むことで、本研究における解析対象企業は49社、変数は30となった(使用した変数の一覧を表41に示す)。なお、企業パフォーマンスにはROAを使用し、東証二部製造業社の平均値は3.3395、標準偏差は4.1857である。よって標準偏差一つ分以上は7.5253、一つ分以下は-0.8462である。

7.5.3 解析手法

作成した解析用データは、反復潜在学習法 (Repeated Potentiality Assimilation, RPA) を用いて解析を行う。

当解析ではサイクル数の上限を 10 回とし、また、ネットワークに投入するデータは、モデル学習用に 49 社の約 70%にあたる 35 社を用い、残りをモデルの汎化誤差調査用データとした。くわえてこの 49 社からなるモデル学習用データは使用するサンプルをランダムに変更され、10 パターンのデータが作成される。これはすなわち、10 個のモデルが作成される意味である。10 パターンのデータを作成するのは、モデル学習時に偶然学習に最適なサンプルからなるデータが使用されるのを防ぐ目的である。

7.5.4 解析手順

これまで解析対象データ、解析データ作成、解析手法について述べてきたが、ここでこれまでに述べた解析手順についての整理と今後の手順について述べる。本研究において解析は、

- 東京証券取引所市場第 2 部上場の製造業企業を対象に社是を収集する (214 社)。
- 収集した社是に対して形態素解析を実施する。
- 「品詞」と「単語カテゴリ」、そして「単語ドメイン」の三つに対して出現回数 (TF) を集計 (内容語 (「名詞」、「形容詞」、「動詞」) を対象とした)、この TF に出現具合を加味した重み付けを行い (TF-IDF)、正規化を実施する (数値化された社是データの作成)。
- 数値化された社是データから解析対象企業を絞り込む (解析データの作成)。
- 解析データ¹¹を RPA により解析し、社是を入力として企業パフォーマンス (総資産利益率, Return On Assets, ROA) のクラス (高い・

¹¹データは 0 から 1 の範囲に標準化され解析に使用された。これは x' を標準化後の値とすると、 $x' = (x - \min(x)) / (\max(x) - \min(x))$ で求める手法である。

低い) を出力するモデルを作成する。

- 作成されたモデルの精度を確認する。
- モデルを解釈し、どの入力変数が出力に影響を与えているか調査する。

の手順で実施される。

7.5.5 結果と考察

7.5.5.1 社是分類モデル

RPA を用いて社是を解析した結果を図 85 に示す。前述したように、RPA は情報量を高めるために反復して計算を行う。この反復回数と情報量、そして汎化誤差の関係を示したものが図 85 である。

図 85 (a) より、反復回数 (サイクル数) が増加するにしたがい情報量が高まっていることが確認できる。一方、図 85 (b) で示すように、汎化誤差は情報量が高まるに従い誤差が増加する結果となった。このことは情報量増加により特定のニューロンのみが活動するようになったものの、データが複雑で小数のニューロンでは対応できなくなっていることを示唆している。

図 86 に各サイクル数での入力層-中間層間の重みを示している (サンプリング回数 10 回目の結果)。図中の四角形が重みを表しており、黒色が重みの符号がプラスであることを、白抜きが重みの符号がマイナスであることを、そして図形の大きさが重みの値の大きさを示している。図 86 からサイクル数が増加するに従い活動しているニューロンが絞られている様子が確認できる。そして図 86 (j) で示されるサイクル数 10 の時点では二つの中間ニューロンが活動するのみである。

図 85 (b) より最も汎化誤差が小さくなるのはサイクル数が 3 の時点であった。よって、社是の内容によりその社是を持つ企業のパフォーマンスが好調な企業とそうでない企業を分類する誤判定率 0.4143 のモデルが作成できたことが確認された。精度比較のために同一データを従来のニュー

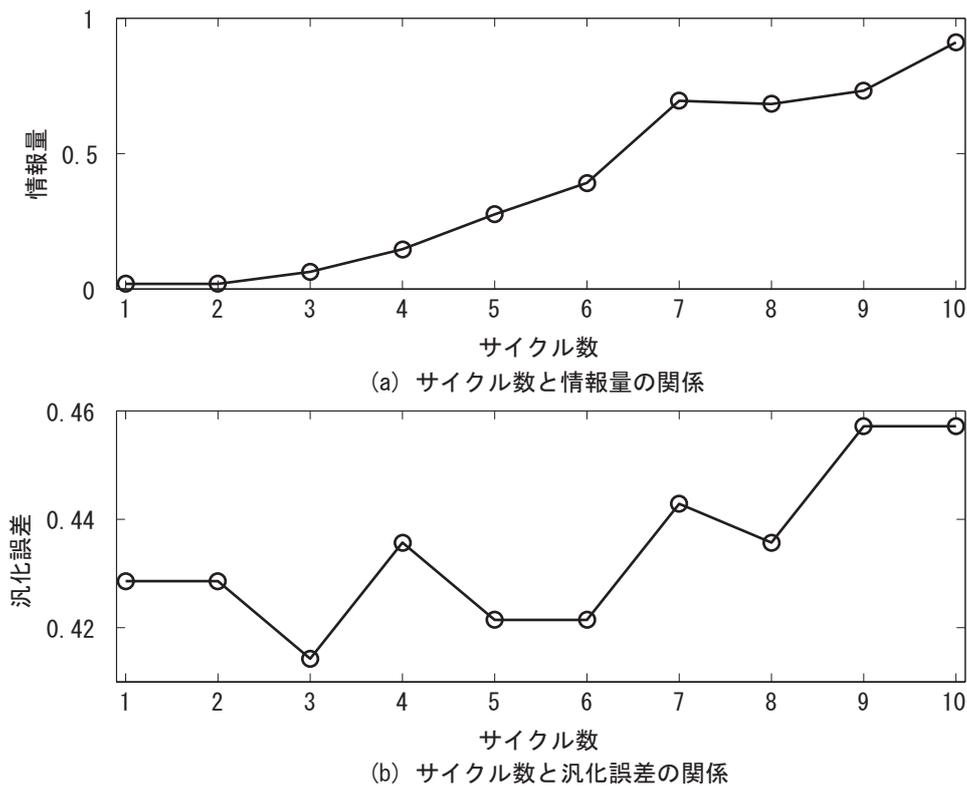


図 85: 情報量と汎化誤差

ラルネットワーク (誤差逆伝播法) を用いてモデル作成したが、こちらは誤判定率が0.4643であり、これによりRPAの持つ情報量最大化効果が発揮されたことが確認できた。

7.5.5.2 モデル解釈

続いて作成したモデルについて、重要変数の確認と解釈を実施した。図85 (a) で確認できるようにサイクル数3の時点の情報量は高い状態ではない。よって多くのニューロンが活動している状態である (図86 (c) を参照)。

解析の流れでも述べたが本研究でモデルは10個作成されているので、図86 (c) に示すような重みが10個得られている。モデルを解釈するにあたり、この10個の重みの平均を求め、そして最も大きい重みを持つものを解釈対象とした (図87 (a) に重みを示す)。また、中間層-出力層間の重みも同様に平均を計算し、入力層-出力層間で最も大きい重みを持った中間ニューロンからの重みを用いた (図87 (b) に重みを示す)。

図87 (a) に示す重みを降順に三つ抽出し解釈を行った。最も大きい重

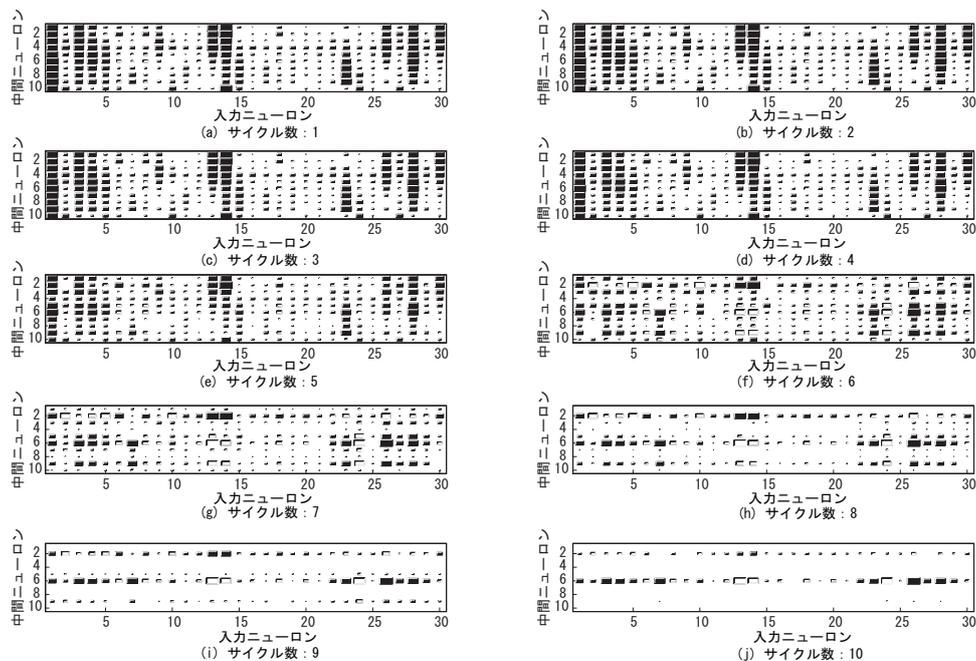


図 86: サイクル数による重み (入力層-中間層) の変化

みは1番目の入力ニューロンから接続されている中間層ニューロンであり、続いて14番目、そして28番目という順番であった。これらは入力した変数のうちで最も出力に影響を与えている変数である。

図 87 (b) で示すように中間層-出力層間の重みは第1出力ニューロンへマイナスの重み、第2出力ニューロンへプラスの重みで接続されていることより、重要とされた変数の値が大きいほうが第2出力ニューロンが活動することとなり、すなわちパフォーマンスがよい企業であることを示す (当解析では企業パフォーマンスが高い企業を「01」、そうでない企業を「10」の2値で表現しており、2ビット目が第1出力ニューロンと、1ビット目が第2出力ニューロンと対応している)。

1番目の変数「名詞」、14番目の変数は「抽象物に関する単語 (うち最多出現単語は貢献)」、そして28番目の変数は「ビジネスに関する単語 (うち最多出現単語は客)」であった。

抽出できた変数の「貢献」「客」について共起関係にある単語について調べたところ「貢献」はその単語の前に「発展 (名詞)」「社会 (名詞)」、その単語の後に「する (動詞)」「こと (名詞)」、「客」はその単語の前に「図る (名詞)」「品質 (名詞)」その単語の後に「ニーズ (名詞)」「タイム

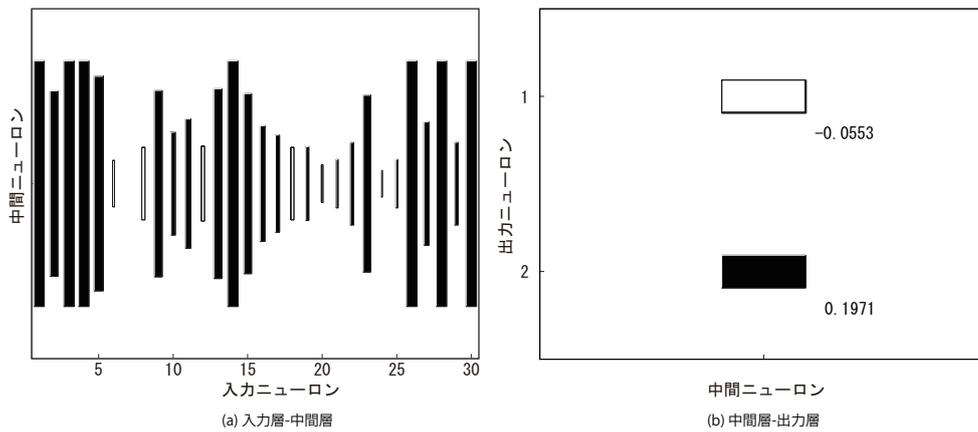


図 87: サイクル数が3時点の重み (中間ニューロン毎の平均値の合計で最大値を持つニューロンの平均重み)

リーだ (形容詞)」という単語と共起関係にあることが分かった。

これより、企業パフォーマンスが好調な企業の社是には「発展に貢献すること」や「社会貢献する」といった目標、または、「客のニーズ」といった具体的な対象に対する目標が記載されている傾向にあると判断できる。

これらの単語は対数の底を2として、

$$MI = \log_2 \frac{\text{共起回数}}{\text{共起語の期待値}}$$

から求められる MI 値を計算し [101] 共起状態を調査した。この値は 1.58 を越えると有意に共起関係があるという意味を表す。先に述べた「貢献」あるいは「客」と共起関係にある単語はどれも $MI > 1.58$ を満たしており有意な結果である。

7.5.6 まとめ

本論文は東証二部上場の製造業社を対象に、社是とパフォーマンスの関係について社是に含まれている単語を軸に論じたものである。企業パフォーマンスには RAO を使用し、東証二部製造業社の平均 ROA (3.3395) よりも標準偏差分大きな ROA を持つ企業を企業パフォーマンスが好調である企業とし、標準偏差分よりも小さな ROA を持つ企業を好調ではない

企業として、どのような社是を持つ企業のパフォーマンスがよいのかを調査した。

49社の企業が解析対象となり、情報量理論に基づいて情報量を最大化させるニューラルネットワークを用いて解析したところ、汎化誤差0.4143で社是から企業のパフォーマンスを予測するモデルが作成できた。作成されたモデルを解釈したところ、「名詞」が品詞のなかで最も重要であり、続いて「抽象物に関する単語」、そして「ビジネスに関する単語」が重要であるという結果となった。

このうち抽象物に関する単語では「貢献」が最頻単語であり、ビジネスに関する単語では「客」が最頻単語であった。これらの単語の共起関係を調べたところ貢献は「発展」「社会」「する」といった単語と、客では「ニーズ」といった単語と共起関係にあることがわかった。

当解析はこれまでに実施した社是と企業パフォーマンスとの関係についてデータと解釈方法を改良したものである。これまでも「社会」という単語の抽出には成功しており「社会貢献」が重要なキーワードであると結論付けていたが、TF-IDFを使用し、共起語を解釈することにより、より詳しく社是の構造を把握することが可能となった。特に客と共に出現する「ニーズ」が抽出できたことより、企業パフォーマンスのよい企業は「世の中に対して訴求する社是を持つ企業」と「企業の主顧客に対して訴求する社是を持つ企業」をもっていることが明らかになった。社是の訴求対象としては他に「従業員」も考えられる。また、当解析で得られた結果は小田らによる研究結果と異なるものであった。今回の手法を他の業種や他の上場部の企業に適用し、さらに深く社是とパフォーマンスの関係を研究していきたい。

8 各解析手法まとめ

本論文では4章、5章、6章、7章にて「消費者来店動機解析」、「履修科目決定行動解析」、「重要 Tweet 判断基準抽出」、「企業行動原理抽出」について述べてきた。ロジスティック回帰分析、一般的なニューラルネットワークとの比較を行う形で各解析を実施してきたが、どの解析においても情報論的ニューラルネットワークは高い予測精度・モデル解釈可能であった。この結果より本論文で目標とする「ニューラルネットワークの脱ブラックボックス化」が達成できたと結論づける。

8.1 各解析結果

本論文では大きく分けて四つの解析事例を示したが、これらのうち「消費者来店動機解析」は「関東地域1店舗を対象とした解析」、「北海道地域5店舗を対象とした解析」、「全店舗を対象とした解析」の三つの解析、「重要 Tweet 判断基準抽出」は「Tweet 分類モデル作成検証」、「重要判断基準抽出」の二つの解析、そして「企業行動原理抽出」は「東証二部上場の4業種を対象とする解析」、「東証一部製造業企業を対象とする解析」、「東証二部製造業企業を対象とする解析」の三つの解析からそれぞれ構成されており、合計九つの解析について述べてきた。また、各解析では(東証二部上場の4業種を対象とする解析を除き)マーケティング分析で一般的に使用されているロジスティック回帰分析、従来のニューラルネットワーク(誤差逆伝播法)、本論文の提案手法である各種情報論的ニューラルネットワークを対比させるかたちで解析を行った。

ここで、各解析結果についてまとめ、手法による解析精度・モデル可読性などといった手法による解析結果の違い、本論文で目指したニューラルネットワークの「脱ブラックボックス」について述べる。表42は各手法によるモデルの精度を順位で示したものである。この表より各解析においてどの手法によるモデルが最も精度が高かったのかが確認できる。これによるとBPは精度1位を多く達成しているものの、精度5位も多く達成しておりバラツキが見られることが確認できる。Logi Regも同様に

表 42: 手法別モデル精度一覧 (汎化能力)

解析データ	高精度 ----- 低精度				
	1位	2位	3位	4位	5位
関東地域1店舗を対象とした解析	PL	SVM	SBP	Logi Reg	BP
北海道地域5店舗を対象とした解析	Logi Reg	PL	SBP	BP	
全店舗を対象とした解析	BP	ABP	DL		
履修科目決定行動解析	BP	Logi Reg			
Tweet 分類モデル作成検証	BP	Logi Reg			
重要判断基準抽出	PL	BP	SBP		
東証二部上場の4業種を対象とする解析	GIM				
東証一部製造業企業を対象とする解析	SPL	SBP	BP		
東証二部製造業企業を対象とする解析	RPA	BP			

ロジスティック回帰分析：Logi Reg, サポートベクターマシン：SVM, 誤差逆伝播法：BP
 自己組織化マップによる事前学習併用の誤差逆伝播法：SBP
 潜在学習法：PL, 反復潜在学習法：RPA
 潜在学習法 (SPL)：SPL, 反復情報量最大化法：GIM
 深層学習：DL, 自己符号化器による事前学習併用の誤差逆伝播法：ABP

表 43: 高精度モデルの作成に成功した割合

	BP	Logi Reg	PL	提案手法全体 (PL・RPA・SPL)
適用回数	8	4	3	5
高精度 (1位) となった回数	3	1	2	4
割合	0.3750	0.2500	0.6667	0.8000

バラツキが確認できる。一方、PL (亜種である SPL と発展系である RPA を含む) は適用した解析で多くが精度 1 位であり、バラツキはほぼ生じていないことが確認できる。ここでこの 3 手法について 1 位を達成した割合を確認すると (表 42 に結果を示す) BP は九つの解析事例に対して 8 回適用されており、そのうち 1 位となった回数は 3 回であり約 38% の達成率であった。Logi Reg は 4 回適用され、達成率は 25%、そして PL は 3 回適用され、約 67% という達成率であった。これにより PL が最も高精度を達成したモデルであることが確認できた。この PL は亜種である SPL と発展系である RPA を合算し「提案手法全体」として達成率を算出すると 80% になり、適用したほとんどの解析において最高精度を達成したことが見てとれる。

8.1.1 ロジスティック回帰分析による解析結果まとめ

ロジスティック回帰分析は「履修科目決定行動解析」、「関東地域 1 店舗を対象とした解析」、「北海道地域 5 店舗を対象とした解析」、「Tweet 分

類モデル作成検証」の4解析で適用した。各解析は使用データや規模が異なるため、一律に精度比較はできないが、表42より判断すると「北海道地域5店舗を対象とした解析」のみ最も高精度なモデル作成に成功したということが見て取れる。

一般的にマーケティング分析で使用されている手法が8解析中(東証二部上場の4業種を対象とする解析はGIMのみ使用のためこれを除く)1解析でのみ最も高精度であったというこの結果は、マーケティング分析で対象とするデータに手法が対応できていないといえる結果である。精度が十分ではないモデルが作成された場合に、企業はモデルに満足して使用しているのか、不満を抱きながらも使用しているのか、あるいは他に使用できる解析手法がない(顧客からの解析手法の指定や上司の説得といった問題により)のか、ということまでは判断できないが、この結果よりロジスティック回帰分析による解析は精度面での不足が目立った結果であった。

ただこれはロジスティック回帰分析が統計的手法であるために精度が低いというものではない。統計的手法であるロジスティック回帰分析とニューラルネットワーク手法である単純パーセプトロンは数学的に等価であるからである[102]。では何が精度に影響を与えているかという点、それは処理の複雑さであろう。表42をロジスティック回帰分析と、単純パーセプトロンを発展させた多層パーセプトロン(誤差逆伝播法)に着目して比較すると、消費者来店動機解析を除くとどれもロジスティック回帰分析よりも誤差逆伝播法の方が精度が高いことが見える。よって、多層パーセプトロンで使用されている中間層が精度に大きく影響を与えていると考えられる。逆に多層パーセプトロンよりもロジスティック回帰分析の方が精度が高い場合(消費者来店動機解析で使用したデータの場合)はデータ処理に中間層での処理が必要なほど複雑ではなく、中間層での処理のためにかえって精度が落ちたと推測される。

また、モデルについての評価を精度ではなく可読性という点に移すと、ロジスティック回帰分析より重要変数についての情報を得られるものの解釈ができないという結果であった(「関東地域1店舗を対象とした解析」

と「北海道地域5店舗を対象とした解析」の二つの解析結果を参照のこと)。さらに、「関東地域1店舗を対象とした解析」および「Tweet分類モデル作成検証」の解析では係数を求める計算が終了せず、くわえて「関東地域1店舗を対象とした解析」では一部の変数のみオッズ比が非常に高い値となっており、多重共線性の発生が疑われる。このようなデータを解析し解釈を実施する場合は、ロジスティック回帰分析の適用は最適ではないと言える。

それでは多重共線性を除いたデータを解析すればよいという話になるが、仮説に基づいて収集した変数を解析手法の制約のために削減しなければならないということは（「Tweet分類モデル作成検証」で発生）、解析として疑問を感じざるを得ない。また、機械的に変数を削除する手法もあるがこれは削除の基準がマーケティング仮説上のものでなく手法上の統計的基準で削除するため、この手法は多重共線性の発生したデータの解析や解釈性の問題は解決するものの、マーケティング分析の観点から考察すると適用は最適なものではない。変数削減はマーケター判断による方が好ましいと言えるからである。よって、変数削減が不要な手法と必要な手法があれば、不要な方が解析に利用しやすいマーケティング分析手法として軍配があがるだろう。データ解析は探索的な側面があるため使用変数の入れ替え作業自体は解析に必要であるものの、解析時点で変数を手法の制約上除くことは、問題解決を目的としてデータを解析するのか、望む結果が出る変数の組み合わせ探索を目的としてデータを解析するのかわからないと言える。また、変数の削減は解釈の取りこぼしに繋がる可能性もある。さらに、さまざまな情報源よりデータを収集し実施するマーケティング分析において、多重共線性が生じていないということは極めてまれな事象であり、多重共線性は避けることが容易ではない。このようなことより、データの状態に解析が左右されるロジスティック回帰分析は一般的に使用されているものの、容易に使用できる解析手法ではないと言える。なお、ニューラルネットワーク手法では多重共線性は問題とならない [25] ため、多重共線性が発生しやすいマーケティング分析に適用させやすい手法であると言える。

8.1.2 誤差逆伝播法による解析結果まとめ

誤差逆伝播法は8解析で使用されモデルの精度が評価された。その結果「関東地域1店舗を対象とした解析」、「北海道地域5店舗を対象とした解析」、「東証一部製造業企業を対象とする解析」、「東証二部製造業企業を対象とする解析」の四つの解析で最もモデル精度が低い結果であった。しかし、「履修科目決定行動解析」、「全店舗を対象とした解析」、「Tweet分類モデル作成検証」、「重要判断基準抽出」の4解析では他手法よりも精度が高く、必ずしも誤差逆伝播法が精度が低いというものではなく、データに左右されることが見て取れる。

また評価基準を精度ではなくモデル可読性に変更すると、解釈は難しいという結果であった（「全店舗を対象とした解析」図 67 および「東証二部製造業企業を対象とする解析」図 86 (a) 参照のこと）。これらの重みを示した図より明らかであるが、誤差逆伝播法の重みは多くのニューロンでそれぞれの値を持つため、解釈を特定のニューロンに絞ることができず、解釈が困難であるという特徴がある（ブラックボックス問題）。今回の結果より改めてそれを確認する結果となった。

解析ができて解釈ができないという問題は、消費者の行動を知ろうとするマーケティング分析では、解析できていないことと同義であるといっても過言ではない。もちろん予測結果だけを重視して、マーケティングの決断を下すことも可能であるがそれは説得力をもたず、例外的な場合のみに行われるにすぎないだろう。よって、この解釈可能という点を解決しない限りニューラルネットワークがマーケティング分析に活用される可能性は高くない。解釈ができない原因であるが、それはほぼすべてのニューロンが活動を行っていることに起因している。多くのニューロンが活動しているということはそれだけ情報処理を複雑に実施しているということであるが、可読性に目を向けるとこれは何を行っているのか分からないということと表裏一体である。

8.1.3 深層学習による解析結果まとめ

本論文で使用した深層学習は4層構造をしたネットワークである。同じく本論文で使用している誤差逆伝播法が3層構造をしているため、1層分が多いネットワークである。このネットワークに自己符号化器を2層使用してデータから特徴を抽出しようとする試みを「全店舗を対象とした解析」で行った。また、層の増加による効果を測定するために3層構造であり事前学習に自己符号化器を用いるネットワークも適用した。その結果は表 42 に示すとおり、BP、ABP、DL の順で精度が高いという結果であった。これはすなわち深層構造がデータ解析に活かされていないという結果であった。

モデル可読性という観点では ABP と DL の重みにおいて (図 66 と図 65 を参照のこと) 入力変数のグループで傾向が見える結果となり、一方誤差逆伝播法の重み (図 67 を参照のこと) ではこのような傾向は見えない結果であったがこれは自己符号化器による特徴抽出の影響である。よって深層学習ではモデル可読性が誤差逆伝播法よりもあるといえる結果であった。しかし、モデル可読性とモデル精度は両立せず、本論文で目指す「高いモデル精度とモデル可読性の両立」による「ニューラルネットワークの脱ブラックボックス」を満足させる結果ではなかった。

これらに加え、深層学習のデータ解析への応用は例えば層数の決定など多くの課題が存在している。本論文では4層構造を採用してスーパーマーケットデータの解析を実施したが、前述した通り4層構造では高いモデル精度を達成できなかった。しかしこれは5層、6層と層を増加させれば満足のいく精度を得ることができたかも知れず、このように何層構造にすれば最適な精度が得られるのか手探り状態である。さらに、1層で使用するニューロンの数も最適な個数はわからないものとなっている。中間ニューロン数の最適個数は従来のネットワークでも同様に探索的であるが、深層学習は層数が多いために探索的なパラメータ設定項目がさらに多い状態となっているのである。

また、データ解析への深層学習の適用は各層でネットワークが何を学

習しているのかがわからないという問題が存在している。重みを確認することで学習内容を追跡することは可能であるが、それを人間が解釈できるかというのは別問題である。入力データが画像といった視覚的に把握できるものであれば、例えば物体のエッジを学習しているなどと学習内容を理解することができるが、入力データが視覚的に把握できない場合は、深層学習の学習内容が理解できず、多層構造での学習が活きてこない状態となる。

8.1.4 提案手法による解析まとめ

8.1.4.1 潜在学習法

潜在学習法は潜在学習法のバリエーションの一つである SPL を含め、4 解析で使用されモデルの精度が評価された。その結果、「北海道地域 5 店舗を対象とした解析」を除いて全ての解析事例で最も高い精度を示した。なお「北海道地域 5 店舗を対象とした解析」で最高精度を記録したロジスティック回帰分析との精度差は 0.0043 であり、これはほぼ同等の精度である。

モデル可読性は潜在学習法を適用した「関東地域 1 店舗を対象とした解析」、「北海道地域 5 店舗を対象とした解析」、「重要判断基準抽出」、「東証一部製造業企業を対象とする解析」の全てで解釈が可能という結果であった。図 88 は「北海道地域 5 店舗を対象とした解析」を実施した際の重みを BP・PL の手法別に示したものである（データサンプリング回数は 10 回目時点）。これによると誤差逆伝播法と比較して活動しているニューロンが絞られていることが見て取れる。くわえてこの状態であってもモデル精度は高く、高精度とモデル可読性を併せ持った手法であるといえる。ゆえにこの潜在学習法が本論文で目指す「ニューラルネットワークの脱ブラックボックス」を具現化した手法であると結論付ける。

モデル可読性についてここでは特に第 4 章でとりあげた POS データ解析に絞るが、第 4 章の図 55 および図 59 に示す潜在学習法による重要変数抽出結果（「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」傾向にあること、および、「昼食・夕食準備時間帯に青果・加工食品・一般食品を購入する顧客」が重要顧客であること) は、図 56 および図 60 に示すロジスティック回帰分析では得ることのできない結果 (「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、および、「9 時台から 20 時台にさまざまな商品を購入する」購買行動が重要であること) であり、潜在学習法がマーケティング分析に有効であることを示す結果となった。

1.5.2 においてマーケティング分析の結果は意思決定に用いられることを述べたが、この第 4 章の潜在学習法による解釈結果は解析後の意思決定に活用することが可能であり、マーケティング施策策定に活用することが可能である。潜在学習法により得られた解釈結果からは「食事準備時間帯にタイムセールを実施する」、「店舗近隣に住んでいる独身者に向けて自炊をすすめる」、「新しい料理レシピを提案する」など、時間に関する施策と食品に関する施策を組み合わせる複合的な多くの施策を策定することが可能である。この施策案は分析結果から得られたことを利用して策定できる次の一手と言えるものである。しかし、ロジスティック回帰分析による解釈結果からでは「雑貨の商品拡充を図る」や「取扱商品数の拡充を図る」といった単一的な施策しか策定することが困難であった。これは分析結果より得られたことを直接的に利用するもので、目先の一手と言えるものであり、これでは意思決定に活用できる解析結果ということとはできない。なお、第 4 章の解析にて得られた「青果が重要である」という解釈結果であるが、一般消費者が普段野菜を購入している店舗は「スーパーマーケット」であることが農林水産省が 2004 年 11 月に実施した約 1 千人を対象とした「食料品消費モニター調査」により報告されている [103]。さらに、オール日本スーパーマーケット協会 3 代目会長の荒井によるとスーパーマーケットの本質は「食品業」であり [104]、また『「家庭内食提供業」として「ワンストップショッピング (買物の目的や動機に基づいて、それと一緒に買うと便利な商品をそろえること) の原理」によって品揃えする業態である』という [105] ことから、妥当性のある結果が潜在学習法の解析により得られたと判断できる。

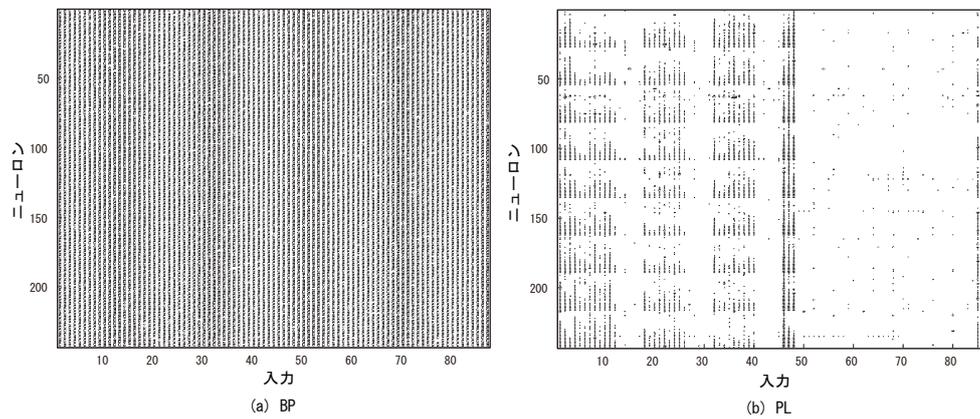


図 88: BP・PL の重み (北海道地域 5 店舗を対象とした解析)

以上により、潜在学習法はロジスティック回帰分析よりも精度が高いモデルが作成でき、またモデル解釈も可能でかつ、その解釈結果はマーケティングの意思決定に活用できることが確認できた。

8.1.4.2 反復潜在学習法

反復学習法は潜在学習法を改良し獲得する情報量を高めたものである。「東証二部製造業企業を対象とする解析」のデータ解析に適用した。この解析では誤差逆伝播法と精度を比較したが、精度は反復潜在学習法の方が高い結果であった。

重みについては図 86 で示したようにサイクル数が増加するにしたがい活動しているニューロンが絞られていることが確認できる。一方誤差逆伝播法では図 89 に示すように各ニューロンが活動しており解釈できないものであった。図には (a) から (j) まで各データサンプリング回数時点の重みが示されているが、どのサンプリング回数時点でも解釈は困難であり、これは誤差逆伝播法の特徴であることが分かる。本論文では図 85 で示したように、最大サイクル数の汎化誤差が最高精度ではなかったため、最高精度であるサイクル数 3 の時点の重みを解析対象とし 10 個モデルの平均を使用して解釈を行った。これは反復潜在学習法を使用することにより活動ニューロンが絞られているために可能な解釈であった。ここで、平均値を用いることにより誤差逆伝播法でも「東証二部製造業企業を対象とする解析」と同様に解釈が可能な印象を受けるが、反復潜在学習法

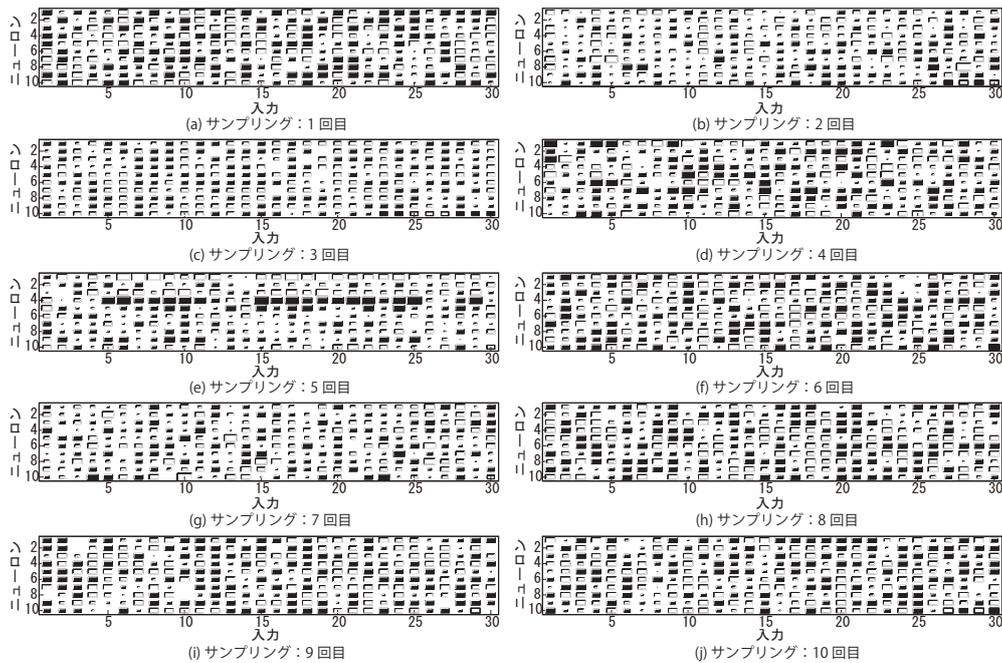


図 89: 誤差逆伝播法の重み (東証二部製造業企業を対象とする解析)

では情報量の上昇により活動ニューロンが絞られるのに対して、誤差逆伝播法では活動ニューロンが絞られることはない。そのため誤差逆伝播法のモデル解釈方法にこれを適用することは適切ではないと考える。誤差逆伝播法に適用すると平均的に大きな値を持つ重みを確認することになるのに対して、「東証二部製造業企業を対象とする解析」での解釈方法はサイクル数の増加に伴い活動ニューロンが減少していくなかで、特定のサイクル数時点で平均的に大きな値を持つ重みを確認するからである。

8.2 提案手法の限界点

ここまで各手法についてまとめを行い、提案手法が本論文で目指した脱ブラックボックス化ニューラルネットワーク手法であると結論を得たが、最後に本論文で提案する情報論的ニューラルネットワークのなかで主提案手法である潜在学習法について限界点について述べる。

潜在学習法は 2.3.3 で述べたように潜在性に着目して学習を実施する。潜在性は分散を基に算出されるものであるので潜在学習法はマーケティング分析的には各自異なる意見を持つ事柄に着目して解析を実施する手法である。マーケティング分析で解析対象となるデータにはパッケージ

表 44: 作成した人工データ

X1	X2	Y	Flg
6.9	2.0	0	1
1.3	3.6	0	0
8.3	9.1	0	1
1.5	1.9	0	0
1.9	3.5	0	0
～略～			
8.4	6.0	1	0
8.7	4.9	1	0
2.5	2.8	1	1
3.7	11.8	1	1
8.4	5.4	1	0

デザインや配色、食品の風味、電化製品の機能など各自で意見が異なる
 いわば表層的なデータと、健康や安全に対する欲求など、意見があまり
 異なる本質的なデータが考えられる。分散に着目して解析を行うと
 はすなわち、表層的なデータのマイニングに特化していることを意味す
 るものであり、潜在学習法はマーケティングデータの解析に向いている
 手法であると言える。

ここで、重要である変数と重要でない変数の分散が同程度である人工
 データを作成し、潜在学習法の変数抽出結果にどのような影響を与える
 のかを確認する。

本検証では表 44 に示すデータを使用した。データの概要は 3 章で使用
 したものと同様である (ノイズ混入率は約 50%)。異なる点は重要変数で
 ある X1 の分散とランダム値より構成されている X2 の分散が同程度 (X1 :
 9.6454, X2 : 9.5754) であることである (図 90 に散布図を示す)。

図 91 はモデルの汎化能力を示したものである。この図よりモデル精度
 が高い順に、潜在学習法 (PL)、重み減衰を使用した誤差逆伝播法 (WD)、
 潜在性を使用しない潜在学習法 (SBP)、誤差逆伝播法 (BP)、変数選択を
 しないロジスティック回帰分析 (Logi Reg ALL)、ロジスティック回帰分
 析 (Logi Reg) であることが確認できる。

ここで自己組織化マップの結果を確認する。図 93 は潜在性を用いない

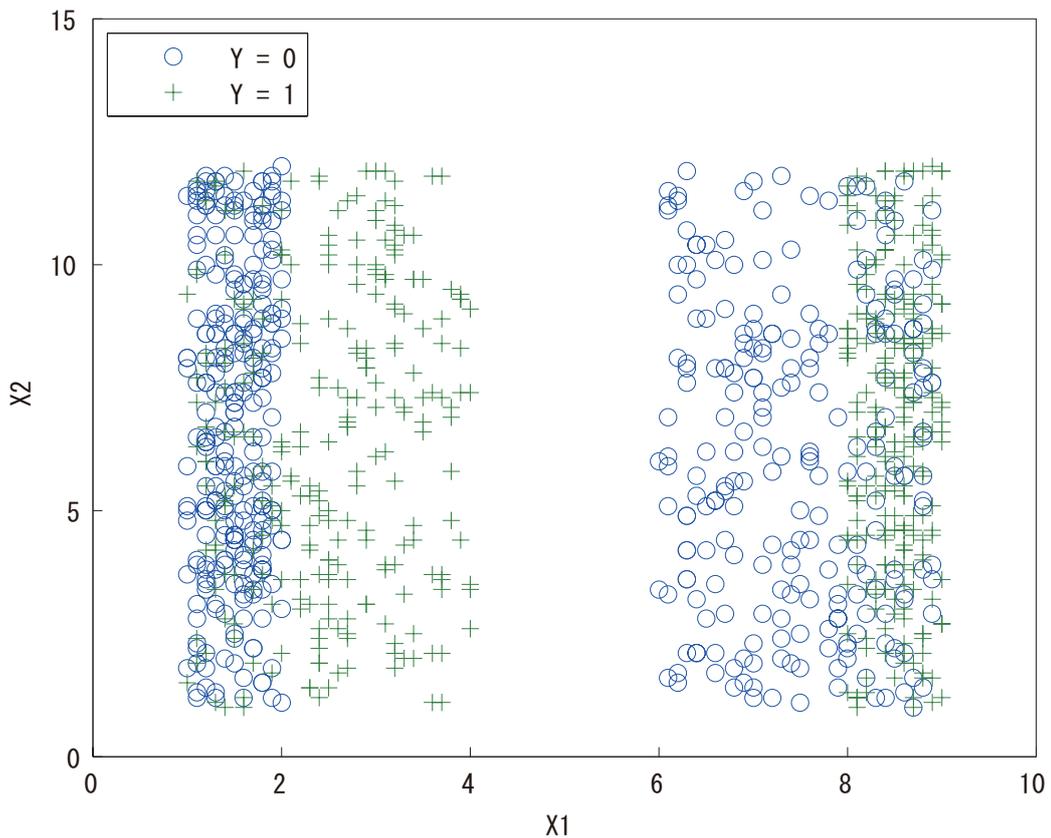


図 90: 作成した人工データの散布図 (同程度分散データ)

潜在学習法での分類結果である。マップの中心部分に距離の離れた箇所が確認でき、そしてマップの右上から左下にかけて境界線が存在していることが確認できる。一方、図 94 は潜在学習法での分類結果である、こちらも境界線がマップの右上から左下にかけて存在していることが確認できる。しかし、図 93 と図 94 の結果は若干異なる部分は確認できるものの、大きな違いは確認できなかった。これは潜在性がデータ分割に大きく働いていない (分散が同程度であるので、どちらの変数も同じくらい分類に活用する。すなわち潜在性を使用していない状態に近い) ためであると考えられる。

続いて変数の抽出について確認を行う。図 92 は入力層-中間層間の重み (図 92 (a) から (c))、潜在性 (図 92 (d))、ロジスティック回帰分析の係数 (図 92 (e) と (f)) を示している。これら図 92 の各図は 10 個のモデル中で最も良い汎化能力を示した際の結果を示している (誤差逆伝播法 : 0.1267、重み減衰を用いた誤差逆伝播法 : 0.1267 (パラメータ $\lambda = 0.9$)、

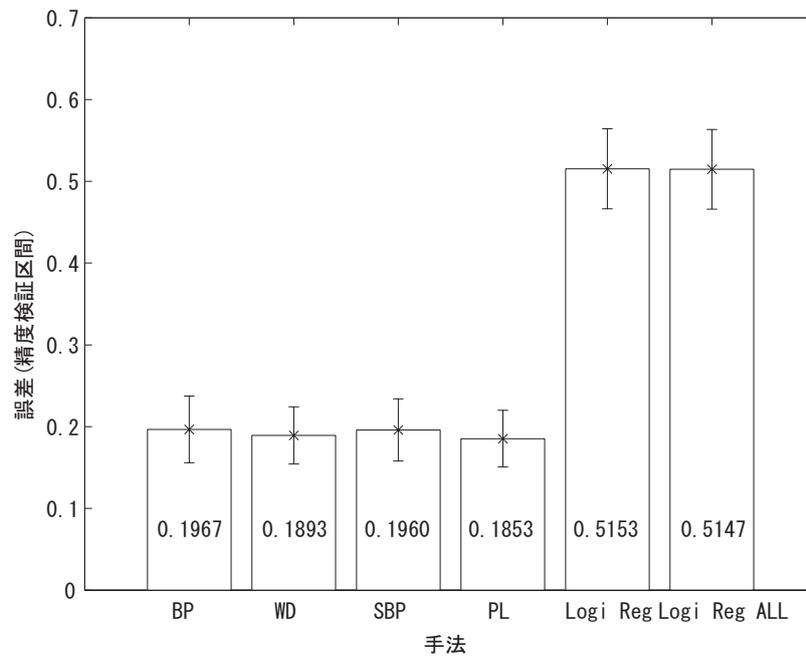


図 91: モデル汎化能力 (同程度分散データ)

潜在性を用いない潜在学習法 : 0.1200、潜在学習法 : 0.1200 (パラメータ $r = 1.4$)、ロジスティック回帰分析 (変数選択あり) : 0.5400、ロジスティック回帰分析 (変数選択なし) : 0.5400)。図 92 の各結果より

誤差逆伝播法

BP の重みは 1 番目の入力ニューロンと接続されている重みの一部で大きな値を持っているものの、それ以外では 1 番目も 2 番目も同程度の重みであり、ネットワーク中で活用されている変数の傾向などを確認することは困難であった。

重み減衰を用いた誤差逆伝播法

WD では重み減衰により重みを持つニューロンが絞られていることが確認できる。1 番目の入力と接続されている重みが大きい値を持っており、また、それらと比較して 2 番目の入力の重みは小さいため、1 番目の変数がネットワークで活用されている傾向にあることが確認できた。

潜在性を用いない潜在学習法

SBP の重みは 1 番目の入力ニューロンと接続されている重みの一部

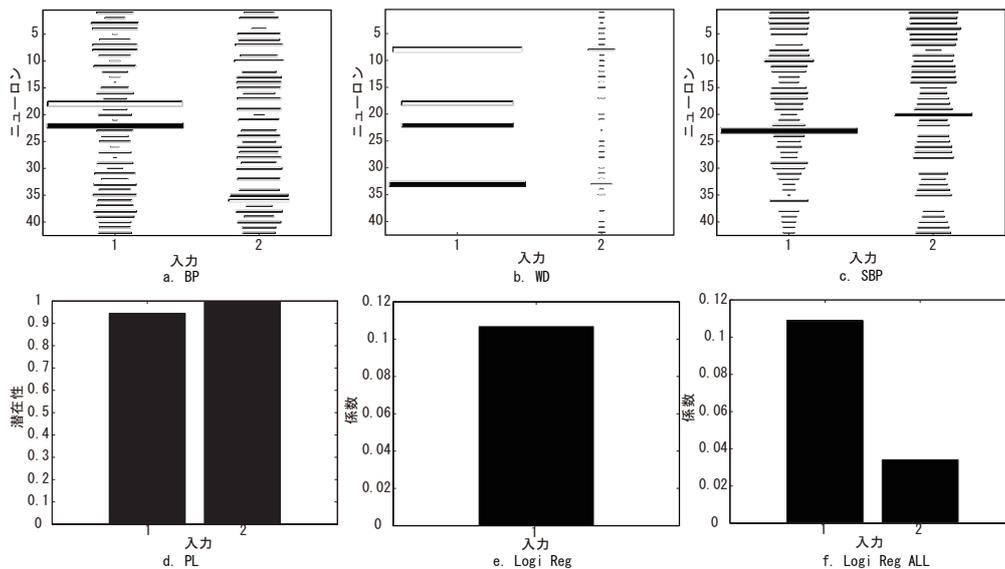


図 92: 重み (入力層-中間層) (同程度分散データ)

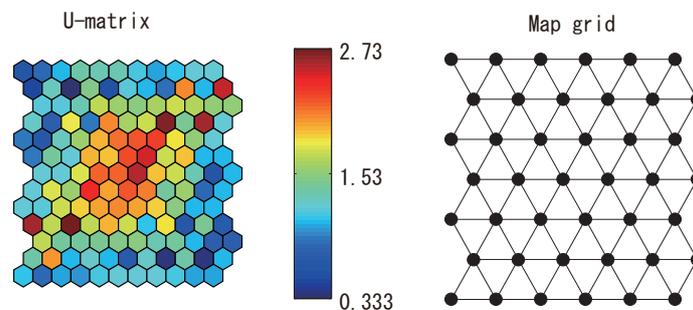


図 93: SBP における自己組織化マップの学習結果 (同程度分散データ)

で大きな値を持っているものの、それ以外では1番目も2番目も同程度の重みであり、ネットワーク中で活用されている変数の傾向などを確認することは困難であった。

潜在学習法

PLでは2番目の入力が1番目の入力よりも大きな潜在性を持っており、よって、2番目の入力が重要であるという解釈となった。

ロジスティック回帰分析 (変数選択あり)

Logi Regの結果であるが、変数削減により1番目の入力のみが変数として採用された。したがって重要変数は1番目変数であると解釈できた。

ロジスティック回帰分析 (変数選択なし)

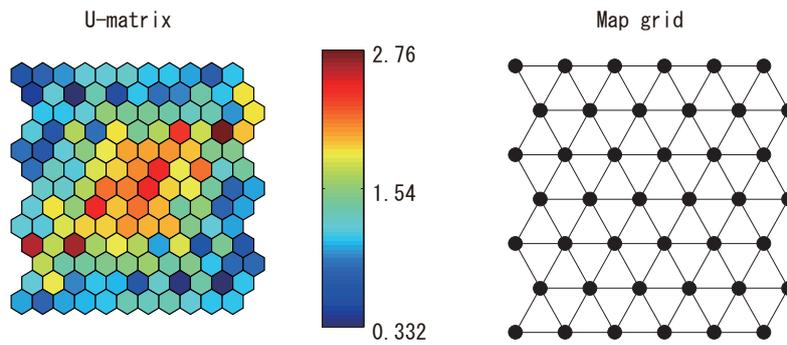


図 94: PL における自己組織化マップの学習結果 (同程度分散データ)

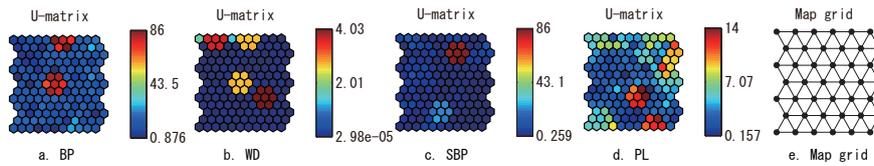


図 95: 学習完了後の重み (入力層-中間層間) を U-Matrix で示したもの (同程度分散データ)

Logi Reg ALL の結果であるが、1 番目の入力の方が 2 番目の入力と比べて大きな係数を持っており、1 番目の入力重要であると解釈できた。

という結果を得た。すなわち正解重要変数を抽出できた解析手法は重み減衰を用いた誤差逆伝播法、変数選択あり、および、変数選択なしのロジスティック回帰分析であったことが確認できた。

さらに続いて U-Matrix にて学習後の重みを確認する (図 95 に結果を示す)。(a) で示す誤差逆伝播法は特定の重みが大きな値を持っていることが確認できる。(b) に示す重み減衰を用いた誤差逆伝播法でも、(a) と同様に特定のニューロンの重みだけが大きい値を持っていることが確認できた。また、(c) の自己組織化マップを事前学習に用いた誤差逆伝播法でも特定のニューロンの重みだけが大きい値を持っていることが確認できるが (a) および (b) の大きな値を持っているニューロンの場所とは異なる場所のニューロンが大きな重みを持っていることが確認できる。最後に (d) に示す潜在学習法の重みであるが、こちらは (a)、(b) そして (c) と比較すると多くのニューロンが大きい値を持っていることが確認できる。濃い暖色の箇所は一部のみであるが、寒色よりも暖色である領域も

多く確認される状態である。この結果は図 92 (d) に示すように潜在性が X1、X2 の両変数に対して大きい、すなわち両変数に着目するという結果から生じたものである。二つの変数に着目して学習をするため複雑な重みになったと考えられる。しかし、図 93 で確認したように汎化誤差は潜在学習法が最も小さいものであり、また図 95 (d) 中の濃い暖色になっている場所を (a)、(b)、(c) と比較すると、場所はこの三つとも違う場所であることが確認できる。潜在学習法は重要変数の解釈という点、重みが複雑であるという点はあるものの、Y を高精度に予測するために必要なニューロンについてはおさえている結果であった。この結果は、X2 には及ばないものの X1 も X2 と同じくらい潜在性を持っており学習時に活用されているために得られたと考えられる。

最後にパラメータ値と潜在性の関係であるが、表 45 に結果を示している。これによるとパラメータ値が増加するにしたがい、変数 1 (X1) の潜在性は増加していること、反面、変数 2 (X2) の潜在性は減少していることが確認できる。そしてパラメータ値が 1.6 の際に潜在性の値は変数 1 と変数 2 で逆転し、以降は変数 1 の方が重要であるという状態になっていることが確認できる。前述した変数解釈では潜在学習法は正解変数の抽出ができていない結果であったが、それはパラメータ r が 1.4 の時点であったためであることが確認できた。重み解釈は最も汎化能力が良い際を選んで実施しているため (図 96 に各パラメータ値と汎化能力の関係を示す) $r = 1.4$ を採用したが、もう少し高いパラメータ値を選択していれば正解変数の抽出に成功したことになる。

しかしモデル解釈には何かの基準が必要であり、これは「最も解釈し易い時点」では恣意的となってしまうため、当検証で用いたように「最も汎化能力が良い時点」を採用する方が好ましく、これにより当解析で使用したようなデータを潜在学習法にて解析すると場合によっては解釈の結果に影響を及ぼすことが確認できた。すなわちこの種のデータ解析は潜在学習法の限界点であると言える。

表 45: パラメータ r の変化と各変数の潜在性 (同程度分散データ)

パラメータ r	X1	X2
1.0	0.9171	1.0000
1.1	0.9252	1.0000
1.2	0.9331	1.0000
1.3	0.9433	1.0000
1.4	0.9449	1.0000
1.5	0.9648	1.0000
1.6	1.0000	0.9981
1.7	1.0000	0.9799
1.8	1.0000	0.9811
1.9	1.0000	0.9657
2.0	1.0000	0.9554

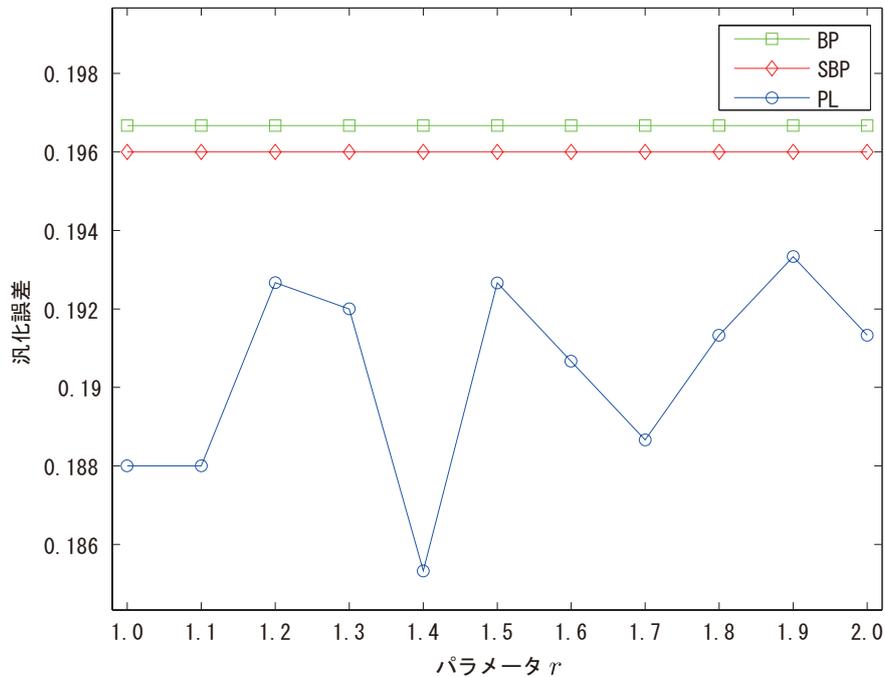


図 96: パラメータ r の変化と汎化能力 (同程度分散データ)

以上の解析結果より、潜在学習法は分散に着目して学習しているため本質的なデータが解析対象であった場合、正しい変数抽出情報を得ることができない場合があることが示された。場合があると述べたのは、潜在学習法がデータの分散ではなく学習第1段階の出力ニューロンの重みの分散を基に学習を行うためである。この出力ニューロンの重みの状態に潜在性算出ロジックは左右されるため「変数抽出が正しく行われる場

合」、「間違ってしまう場合」が発生する。潜在学習法の精度は高精度であることが示されたものの、マーケティング分析では解釈が重要であるため、当解析のように各変数の分散が同等のデータを解析することは潜在学習法は得意としておらず、手法の限界点であると言える。

実務的な観点から見れば、本質的なデータは「意見が異なっていない、すなわち、解析するまでもなく大切」ということになるため、解析データとして実際のデータ中に存在することは多くないと想定されるものであるが、マーケティング解析対象はさまざまであるため、対応できない場合があるデータの存在は改善すべき事柄であり、今後は分散以外の着目点も使用するなどの改善を行っていききたい。

9 結論

本論文はニューラルネットワークがマーケティング分析に一般的に選択されない現状を鑑み、その主たる原因がニューラルネットワークの持つブラックボックスであるにとらえ、ニューラルネットワークの脱ブラックボックス化について検証、マーケティング分析への適用可能性について論じたものである。本論文では情報論的ニューラルネットワークを脱ブラックボックス化ニューラルネットワーク手法として提案、各種解析に適用してその有効性を確認した。提案手法は複数あるが主となる手法は潜在学習法であり、この手法を用いて解析を実施したところマーケティング分析で一般的に使用されるロジスティック回帰分析、既存のニューラルネットワーク手法である誤差逆伝播法よりも高精度なモデルの作成、そしてモデルの解釈に成功した。モデル解釈結果は、得られた結果を用いて複数のマーケティング施策案を策定できるなど、マーケティング課題に対する意思決定に活用できるものであり、潜在学習法は複雑なデータを解析対象とするマーケティング分析に有効かつ有益であると結論づけることができた。

9.1 SWOT 分析

本論文はニューラルネットワークがマーケティング分析に一般的に選択されない現状を鑑み、その主たる原因がニューラルネットワークの持つブラックボックスであるにとらえ、ニューラルネットワークの脱ブラックボックス化について検証・マーケティング分析への適用可能性について論じたものである。

マーケティング分析では一般的に多変量解析が使用されているが、これを使用することにより、データ予測や分類といった解析結果にくわえモデルを解釈することで出力に影響を及ぼしている入力変数を抽出することが可能である。マーケティングにとって解析結果は重要である。これはマーケティング諸活動が実施にコストがかかること、そして消費者の印象に対しても影響を与えるものであるからである。したがって、根拠な

き解析結果ではマーケティング施策の策定やクライアントをはじめとする関係者の説得が困難である。

ニューラルネットワークは脳での情報処理を模したものであり、高いデータ解析能力を有している手法である。マーケティングの解析対象は人間であるので、人間の情報処理を模した手法でマーケティング分析を実施することは親和性が高いといえるが、マーケティング分析に対して弱点といえる特性を持っている。それは解釈困難という問題である。いくら解析精度が高くても、その結果を出力するに至った過程がわからないと、そのモデルは信憑性に難があり、これではその解析結果がマーケティングプランとして実施される可能性は低い。しかし、ニューラルネットワークのデータ解析能力には定評があるため、この能力は維持しつつ過程の解釈が可能となれば、マーケティング分析で活用される手法となると考えられる。

そこで本論文では既存のニューラルネットワーク手法が持つ解釈困難問題を解決するため、情報論的ニューラルネットワークの使用を提案した。本論文では数種類の情報論的ニューラルネットワークを取り上げているが、主な提案手法は潜在学習法である。これは潜在性に着目して学習を行う新しいニューラルネットワーク学習法であり、モデル解釈が可能となっている。手法の解析特性などは前章までで述べているのでそちらを参照されたい。ここでは、手法についてマーケティング的な観点より全体的に整理を行い提案手法の結論を整理したい。

図 97 は潜在学習法について SWOT 分析を実施した結果である。SWOT 分析 [68], [106] の「SWOT」とは「Strength」、「Weakness」、「Opportunity」、「Threat」の頭文字を取ったものであり、外部環境と内部環境から解析対象をそれぞれ評価するものである。「Strength」「Weakness」が内部環境の分析であり、「Opportunity」「Threat」が外部環境の分析となる。

SWOT 分析の結果、内部環境的な強みは「ロジックが平易」、「モデル解釈が可能」、「複雑なデータ解析が可能」、「高精度な解釈が可能」、「大規模データ解析も可能」という結果となった。そして外部環境的な強みは「ビッグデータ解析が注目されている」、「機械学習が注目されている」

		プラス要因	マイナス要因
内部環境		<ul style="list-style-type: none"> ・ロジックが平易である ・モデル解釈が可能である ・複雑なデータが解析可能である ・高精度な解析が可能である ・大規模データ解析も可能である 	<ul style="list-style-type: none"> ・潜在性の活用が限定的である
		Strength	Weakness
外部環境		<ul style="list-style-type: none"> ・ビッグデータ解析が注目されている ・機械学習が注目されている 	<ul style="list-style-type: none"> ・手法の知名度が低い ・ニューラルネットワークは「解釈が困難」というイメージ
		Opportunity	Threat
		<ul style="list-style-type: none"> ・質的調査が増加している 	

図 97: SWOT 分析 (潜在学習法)

となった。内部環境であげた強みを有していても昨今の解析データ規模が増加している状況に対応できないということになれば、その手法は活用できないということになるが、潜在学習法は「大規模データの解析」も行えるため、近年のデータ状況に対応できるものである。従って外部環境の状態を手法に対する追い風と捉えることが可能であり、潜在学習法は今後マーケティング分析で活用される可能性を持っていると言える。

内部環境的な弱みであるが、「潜在性の活用が限定的」という弱みを有している。これは強みである「ロジックが平易である」こと故に発生している弱みである。潜在学習法は学習の初めに潜在性を加味して学習が行われる。この潜在性は L1 正則化や L2 正則化のように学習ロジックの深層に関わらないためロジックは平易であり、第三者に対する手法説明や他手法への展開に利点がある。しかし、潜在性がロジックの表層で活用されるために限定的な活用となってしまふのである。潜在性の活用については反復潜在学習法にて解消を試みているが、解析適用事例数が十

分ではなくマーケティング分析手法として有効かどうか引き続き検証を続けていく。

外部環境的な弱みは「手法の知名度が低い」、「ニューラルネットワークは『解釈が困難』というイメージ」、「質的調査が増加している」となった。手法の知名度については発表する機会を多くすることで克服は可能と考えられる。しかし、ニューラルネットワークのイメージについては、一般的にイメージされている「解釈が困難」というものは簡単に更新されるものではなく、解釈性が克服されたニューラルネットワークの新しい手法といえども既存イメージによる影響は避けられないと推測される。潜在学習法はニューラルネットワーク手法の一種であるが、実務での普及を目指す際は手法名称を既存のニューラルネットワークのイメージを脱却するようなものとし、解釈が可能な新しい手法であることを示す必要がある。

最後に「質的調査が増加している」という「脅威」であるが、これは近年のデータ解析手法の動向である。図 98 は日本マーケティング・リサーチ協会実施の「経營業務実態調査」の「表 8 調査手法別 売上高構成比」 [107], [108], [109] を示している。「アドホック調査を 100%とした場合」の売上高構成比を引用したが、この実態調査により、どの調査手法がマーケティング分析の業界で活用されているのか動向を把握することができる。この図より質的調査の割合が近年増加していることが確認できる。マーケティング調査手法は数値化されたデータを解析する「量的調査」と、インタビューなどを実施する「質的調査」があるが、近年、質的調査の比率が増してきている状況にあることが図 98 より確認できる。これはデータ解析により得られる結果よりもインタビューなどにより得られる結果の方が重要視されつつあると考えることができる。このことは新しい量的調査手法について提案する本論文には「脅威」である。解釈可能な解析手法 (多変量解析) がデータ解析に主に使用されているにも拘わらず、質的調査に移行していることは、現状のデータ解析手法に限界が感じられてきており調査トレンドが変化しつつあると考えることも可能であるからである。しかし、同時にこれはチャンスと捉えることも

可能である。ここまで示したように既存手法の限界点といえる事柄を多重共線性問題や線形・非線形といった点より考察すると、近年の質的調査の増加は、これらに対応できない既存手法により得られる結果に疑念が生じつつあるのではないかと考えられるからである。本論文の提案手法である情報論的ニューラルネットワークは量的調査の既存解析手法の問題を解消するものであり、既存手法では解析・解釈できなかった結果を得ることに成功したことを前章までで示してきた。よって潜在学習法を使用することで既存の手法からは得ることが困難であった新しい発見につながる可能性もあり、これにより量的調査から離れつつある動向の回復が期待できると考える。このように、近年の調査手法の動向は脅威ではあるもののチャンスになる可能性でもあると言える。

9.2 潜在学習法のポジショニング

ここで、各解析手法について整理を行いたい。図 99 は一般的な多変量解析手法とニューラルネットワーク手法、そして本論文の提案手法について「解析データの複雑さ」と「解釈性」の 2 軸により関係を示したポジショニングマップである。ポジショニングマップは定めた軸を基に相対的な比較を行うフレームワークである。ロジスティック回帰分析 (多変量解析) は解析できるデータの複雑さは大きくないものの、解釈性は高い手法である。そして誤差逆伝播法 (ニューラルネットワーク) は解析できるデータの複雑さは大きいものの、解釈性は低い手法であることが図より確認できる。この二つの既存手法に対して本論文で提案する手法は解釈が容易であり複雑なデータ解析ができるもの、すなわち誤差逆伝播法を「解釈が容易」な方向へシフトさせた位置に位置づけられる手法である。

本論文では九つの解析事例に対しロジスティック回帰分析、ニューラルネットワーク、情報論的ニューラルネットワークをそれぞれ適用して特性を評価した。その結果、手法がロジスティック回帰分析 (多変量解析) であるからといって必ずしもモデル解釈ができるわけではないこと、解析

結果がデータに敏感 (多重共線性データは除くなど解析対象データを選ぶ必要があること) であることが確認された。解析結果が解釈できない問題と、データに敏感であるという問題は、説明力が求められる分析の解析手法としては難があるといえる。特にデータに多重共線性が生じている場合、多変量解析手法の適用は困難である。

一方、ニューラルネットワークによる解析結果では、解釈が困難であるという前評判通り、ブラックボックスは強固なものであり、解釈がほぼできない状態であることを改めて確認した。そこで潜在性という観点からこのブラックボックス問題に挑み、情報論的ニューラルネットワークを適用して解析を実施した。このネットワークは潜在性 (分散) に着目して学習を行うもので、潜在性の観点よりネットワークでの情報処理に必要な情報 (ニューロン) を取捨選択して情報処理を実施する。

第 8 章の表 42 と表 43 に示すように提案手法は本論文で述べたほとんどの解析で最も精度が高いモデルであることを示した。またモデル解釈性であるが、情報論的手法は良好なモデル可読性を示す結果であった。解析は意思決定に活用するために実施するので「解析 → 解釈 → 意思決定」という流れが必要であるが、ロジスティック回帰分析による解析結果では「解釈」から次のステップである「意思決定」へつながる結果を得ることに必ずしも成功せず (POS データ解析結果より判断)、一方、主提案手法である潜在学習法を用いた解析では意思決定へつながる結果の抽出に成功した。

以上で述べたように提案手法は例えば解析の現場で多重共線性データの解析に悩んでいる実務者をサポートできるなど有用な解析手法となると考えられる。また、削減せざるを得なかったデータを解析に活かせるようになることで、既存の手法では明らかにできなかった新しい発見も期待できる。

本論文では分散に着目した情報論的ニューラルネットワークをデータ解析に適用させて精度とモデル可読性を検証し、マーケティング分析に有益な手法であると結論付けた。今後はマーケティングモデルをニューラルネットワークに組み込むなどマーケティング分析により特化した手

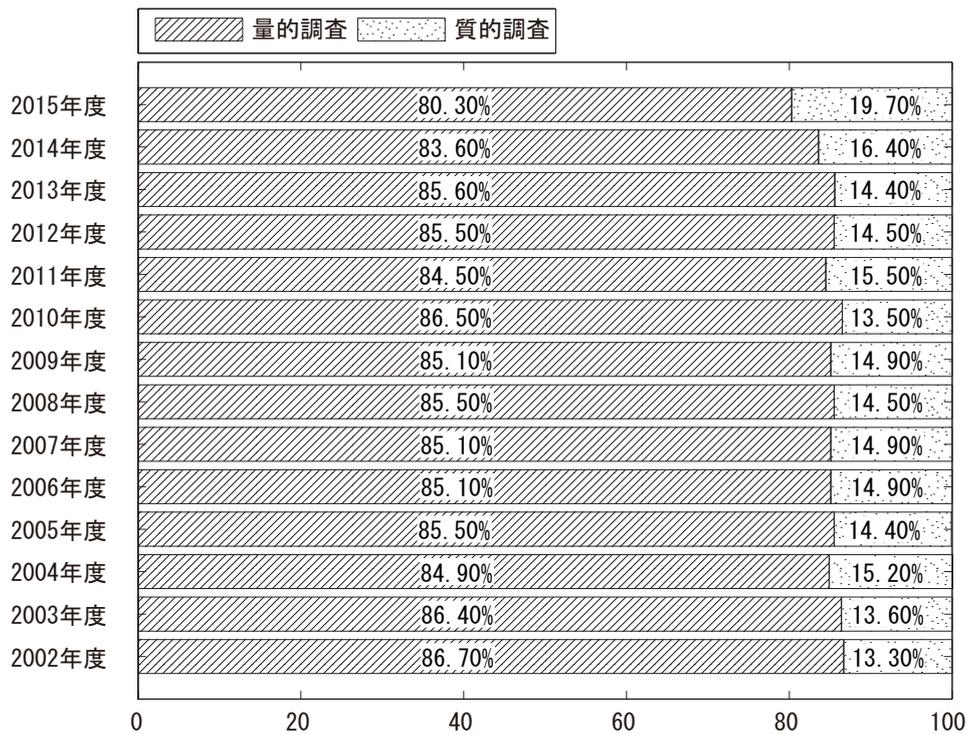


図 98: マーケティング調査手法動向 文献 [107], [108], [109] より引用し作成

法の提案や、分散以外の着目点、SOMの次元数など、必要に応じたネットワーク構成変更を実施し、より有益な解析手法を提案できるよう引き続き研究を行っていく所存である。

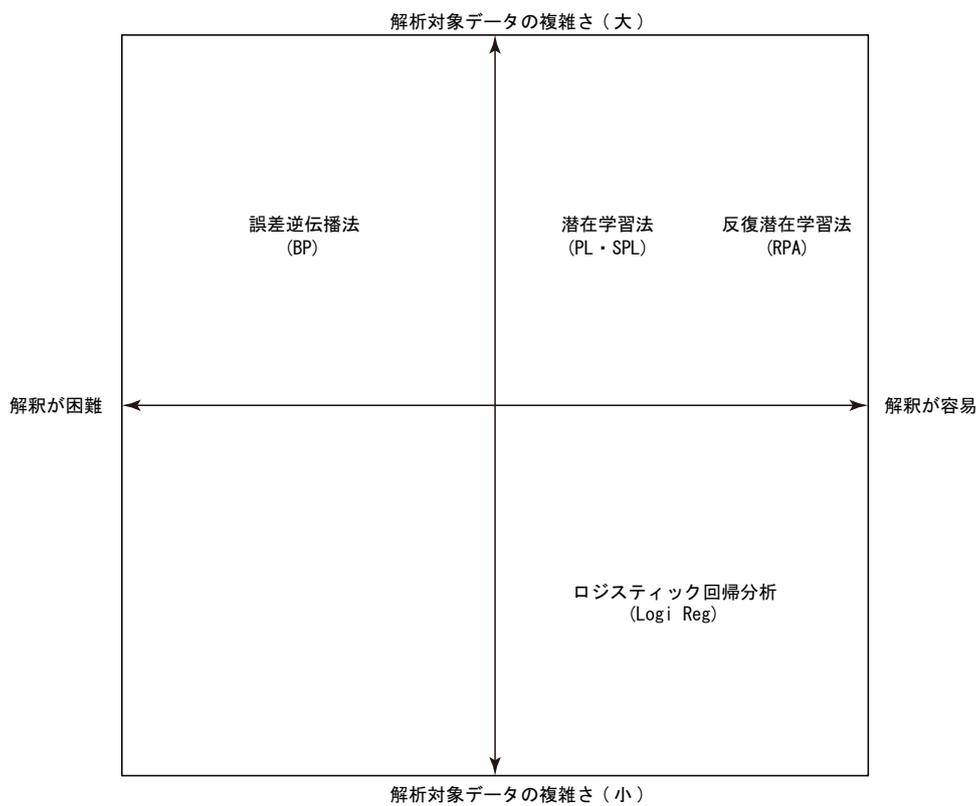


図 99: ポジショニングマップ

謝辞

本論文は、東海大学 上村龍太郎教授、内田理教授のご指導のもとで行った研究内容をまとめたものです。両教授には研究および論文執筆全般に渡り、多大なるご指導を賜りました。

上村教授は博士課程指導教員であります。大学生生活で最初に履修した科目の担当教員でもあり、そのご縁以来、学士課程・修士課程・そして博士課程と、長期また多岐に渡り多大なるご指導を賜りました。

内田教授は博士課程副指導教員であり、博士課程・研究者としての心構えなど多岐に渡り多大なるご指導を賜りました。

また研究指導以外に、両教授の研究に取り組まれる姿勢・お人柄により大変充実した学生生活をおくることができました。ここに深く感謝の意を表します。

本論文を審査委員の立場より審査して頂きました小泉真人教授は学士課程時代の指導教員であり、博士課程におきましても研究にあたり、多岐に渡りご指導・ご助言を賜りました。ここに深く感謝の意を表します。

また、東海大学 情報教育センター (湘南キャンパス) の各先生方より、研究はもちろんのこと、公私共にご指導・ご助言を賜りました。ここに深く感謝の意を表します。

最後に私事ではありますが、博士課程まで勉学の機会を与えて頂き、また常に励まし・支えてくれた両親に深く感謝の意を表します。

参考文献

- [1] 日本マーケティング協会, “マーケティング定義”, http://www.jma2-jp.org/main/index.php?option=com_content&view=article&id=42:newsflash-4&catid=3:newsflash
- [2] E. Jerome McCARTHY, 粟屋義純 (監訳), “ベーシック・マーケティング (初版第2刷)”, 東京教学社, 1979.
- [3] 池尾恭一, “消費者行動とマーケティング戦略 (初版第1刷)”, 千倉書房, 1991.
- [4] 石井淳蔵, 栗木契, 嶋口充輝, 余田拓郎, “ゼミナール マーケティング入門 (初版第1刷)”, 日本経済新聞社, 2004.
- [5] 亀井昭宏, 足田聰 編著, “新広告論”, 日経広告研究所, 2005.
- [6] 和田充夫, 恩蔵直人, 三浦俊彦, “マーケティング戦略 [新版] (新版第11刷)” 有斐閣, 2004.
- [7] 井上崇通, “消費者行動論 (初版第1刷)” 同文館出版, 2012.
- [8] グロービス経営大学院 (編著), “グロービス MBA マーケティング [改訂3版] (改訂3版第6刷)”, ダイヤモンド社, 2013.
- [9] 合原一幸, “ニューラルコンピュータ 脳と神経に学ぶ (初版第5刷)”, 東京電機大学出版局, 1989.
- [10] 竹田辰興, “ニューラルネットワークの応用と今後の発展”, プラズマ・核融合学会誌, Vol. 82, No. 5, pp. 275-281, 2006.
- [11] 田中昭雄, 小峯裕己, 関庸一, 三宅 義彦, “ニューラルネットワークによる住宅エネルギー消費構造モデルの研究”, 空気調和・衛生工学会論文集, No. 111, pp. 9-17, 2006.
- [12] V. M. Souza, L. E. A. Vieira, C. Medeiros, L. A. Da Silva, L. R. Alves, D. Koga, D. G. Sibeck, B. M. Walsh, S. G. Kanekal, P. R. Jauer,

- M. Rockenbach, A. Dal Lago, M. V. D. Silveira, J. P. Marchezi, O. Mendes, W. D. Gonzalez and D. N. Baker, “A neural network approach for identifying particle pitch angle distributions in Van Allen Probes data”, AGU Space Weather Journal, Vol. 14, pp. 275-284, 2016.
- [13] 中田（黒田）祥子, “ニューラルネットワークアプローチによる経済分析（2）－通貨当局の外為市場介入への応用例－”, 金融研究, Vol. 17, No. 6, pp. 99-122, 1998.
- [14] 西尾チヅル, “ニューラルネットワークによるマーケティングモデル”, オペレーションズ・リサーチ, Vol. 39, No. 4 pp. 31-36, 1994.
- [15] 松田芳雄, “顧客分析とデータマイニングの動向”, ユニシス技報, 通巻 68 号, pp. 177-196, 2001.
- [16] 野口博司, “マーケティングにおけるビッグデータの活用について”, 流通科学大学論集-流通・経営編-, Vo. 27, No. 2, pp. 219-235, 2015.
- [17] 鷺尾隆, “データマイニングと学習”, 計測と制御, Vo. 42, No. 6, pp. 480-484, 2003.
- [18] 岩崎学, “データマイニングと知識発見 -統計学の視点から-”, 行動計量学, Vol. 26, No. 1, pp. 46-58, 1999.
- [19] 宮崎和光, 小林重信, “Profit Sharing の不完全知覚環境下への拡張 : PS-r*の提案と評価”, 人工知能学会論文誌, Vol. 18, No. 5, pp. 286-296, 2003.
- [20] 落合光太郎, “異種混合学習技術を活用した日配品需要予測ソリューション”, NEC 技報, Vol. 68, No. 1, pp. 83-86, 2015.
- [21] 福見稔, “ニューラルネットワークからのルール抽出”, 高速信号処理応用技術学会誌（電子技術）, Vol. 5, No. 3, pp. 3-8, 2002.

- [22] 西岡公一, 木村公昭, 坏雅博, 山藤昌志, 藤原佳代, 河内善弘, 藤田健太, “信用リスクに関する研究”, 三菱総合研究所 所報, No. 40, pp. 34-47, 2002.
- [23] 棚田東作, 荒川正幹, 西村竜一, 船津公人, “ケモメトリックス手法を用いた材料開発支援プログラムの開発”, Journal of Computer Aided Chemistry, Vo. 1, pp. 35-46, 2000.
- [24] 松原望, “松原 望 統計学”, 東京図書 (初版第1刷), 2013.
- [25] 大塚友吉, “脳卒中予後予測における頭部CT所見の有用性について”, リハビリテーション医学, Vo. 40, No. 7, pp. 443-452, 2003.
- [26] Yann Le Cun, John S. Denker and Sata A. Solla, “Optimal Brain Damage”, Proceedings of Neural Information Processing Systems 1989 (NIPS 1989), pp. 598-605, 1990.
- [27] 青山智夫, “ニューラルネットワークによる薬物の構造活性相関に関する研究”, 東北大学博士学位論文, 1992.
- [28] 井須芳美, 長嶋雲兵, 細矢治夫, 青山智夫, “分子の構造活性相関解析のためのニューラルネットワークシミュレータ: Neco (NEural network simulator for structure-activity COrrrelation of molecules) の開発”, Journal of Chemical Software, Vol. 2, No. 2, pp. 76-94, 1994.
- [29] Viktoriya Krakovna and Finale Doshi-Velez, “Increasing the Interpretability of Recurrent Neural Networks Using Hidden Markov Models”, Proceedings of 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), pp. 46-50, 2016 (preprint).
- [30] Tautvydas Cibas, Franoise Fogelman Souli, Patrick Gallinari and Sarunas Raudys, “Variable Selection with Neural Networks”, Neurocomputing, Vol. 12, pp. 223-248, 1996.

- [31] 石川眞澄, “ニューラルネットの忘却付き構造学習”, 日本ファジィ学会誌, Vol. 9, No. 1, pp. 2-9, 1997.
- [32] Masumi Ishikawa, “Structural Learning with Forgetting”, Neural Networks, Vol. 9, No. 3, pp. 509-521, 1996.
- [33] Ralph Linsker, “Self-organization in a perceptual network”, Computer, Vol. 21, pp. 105-117, 1988.
- [34] 東京大学教養学部統計学教室 編, “統計学入門 (初版第 27 刷)”, 東京大学出版会, 2008.
- [35] 山本義朗, 鳥越規央, “統計学序論 (初版第 1 刷)”, 東海大学出版会, 2013.
- [36] 岡谷貴之, “深層学習 (初版第 5 刷)”, 講談社, 2015.
- [37] 豊田秀樹, “非線形多変量解析 -ニューラルネットによるアプローチ- (初版第 4 刷)”, 朝倉書店, 2002.
- [38] David E. Rumelhart, James L. McClelland and the PDP Research Group, 甘利俊一 (監訳), “PDP モデル -認知科学とニューロン回路網の探索- (初版第 2 刷)” 産業図書, 1989.
- [39] Teuvo Kohonen, 徳高平蔵, 大藪又茂, 堀尾恵一, 藤村喜久郎, 大北正昭 (監修), “自己組織化マップ 改訂版 (初版第 1 刷)”, Springer, 2005.
- [40] 今井秀樹, “情報理論 (初版第 33 刷)”, 昭晃堂, 2004.
- [41] Norman Abramson, 宮川洋 (訳), “情報理論入門 (初版第 24 刷)”, 好学社, 2012.
- [42] Ryotaro Kamimura and Ryozo Kitajima, “Gradual Information Maximization in Information Enhancement to Extract Important Input Neurons, Proceedings of the IASTED International Conference Artificial Intelligence and Applications (AIA 2014), pp.386-392, 2014.

- [43] 北島良三, 遠藤啓太, 上村龍太郎, “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”, オペレーションズ・リサーチ, Vol. 61, No. 2, pp. 88-96, 2016.
- [44] 市村直幸, 竹内倶佳, 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, Vol. J77-D2, No. 2, pp. 301-310, 1994.
- [45] 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, Vol. 35, No. 10, pp. 2046-2053, 1994.
- [46] 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), Vol. 124, No. 9, pp. 1874-1880, 2004.
- [47] 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, Vol. 4, No. 3, pp. 89-100, 2005.
- [48] Ryotaro Kamimura and Ryoza Kitajima, “SOM Knowledge Induced Learning with Maximum Information Principle to Improve Multi-layered Neural Networks”, Proceedings of the Sixth IASTED International Conference on Computational Intelligence (CI 2015), pp. 255-261, 2015.
- [49] Ryotaro Kamimura, “Repeated Potentiality Assimilation: Simplifying Learning Procedures by Positive, Independent and Indirect Operation for Improving Generalization and Interpretation”, Proceedings of 2016 International Joint Conference on Neural Networks (IJCNN 2016), pp. 803-810, 2016.

- [50] Ryotaro Kamimura, “Solving the Vanishing Information Problem with Repeated Potential Mutual Information Maximization”, *Neural Information Processing*, Vol. 9950, pp. 442-451, 2016.
- [51] Juha Vesanto, Johan Himberg, Esa Alhoniemi and Juha Parhankangas, “SOM Toolbox for Matlab 5”, Technical report on SOM Toolbox 2.0 (Report A57), SOM Toolbox Team Helsinki University of Technology, 2000.
- [52] Christian Spevak and Richard Polfreman, “Analyzing Auditory Representations for Sound Classification with Self-Organizing Neural Networks”, *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, pp. 1-6, 2000.
- [53] R. Crghino and Y.-S. Park, “Review of the Self-Organizing Map (SOM) approach in water resources: Commentary”, *Environmental Modelling & Software*, Vol. 24, pp. 945-947, 2009.
- [54] A.M. Kalteh, P. Hjorth and R. Berndtsson, “Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application”, *Environmental Modelling & Software*, Vol. 23, pp. 835-845, 2008.
- [55] A. Ultsch and H.P. Siemon, “Kohonen’s Self Organizing Feature Maps for Exploratory Data Analysis”, *Proceedings of International Neural Networks Conference (INNC)*, pp. 305-308, 1990.
- [56] 北島良三, 上村龍太郎, “小売店顧客の継続来店予測モデル作成と継続来店顧客の行動分析 -POS データを入力とするニューラルネットワークを用いた予測モデル-”, *Proceedings of Japan Marketing Academy Conference*, Vol. 4, pp. 266-277, 2015.
- [57] 平成 26 年度データ解析コンペティション配布資料, “データ 1 概要”, http://www.namalab.org/140808entry_file.zip

- [58] 山下遥, 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, Vo. 60, No .2, pp.81-88, 2015.
- [59] 小野田崇, “サポートベクターマシンの概要”, オペレーションズ・リサーチ, Vol. 46, No. 5, pp. 225-230, 2001.
- [60] 小野田崇, “特集 サポートベクターマシン：その仕組みと応用－分類手法の新展開－ 特集にあたって”, オペレーションズ・リサーチ, Vol. 46, No. 5, p. 224, 2001.
- [61] Helsinki University of Technology (現:Aalto University), Laboratory of Computer and Information Science, “SOM Toolbox 2.0”, <http://www.cis.hut.fi/somtoolbox/>
- [62] 西尾チヅル, 宮澤永光, 中西祥八郎, “ニューラルネットワークとファジィ推論を用いたセールスプロモーション効果分析システム”, 日本経営工学会誌, Vol .42, No. 6, pp. 431-439, 1992.
- [63] 渡辺亮, 北村裕人, 星野直人, 関庸一, “買回りタイプによる顧客購買行動の理解”, オペレーションズ・リサーチ, Vo. 50, No. 9, pp. 644-653, 2005.
- [64] 北島良三, 白澤秀剛, 丸山有紀子, 田中真, 坂田圭司, 上村龍太郎, “大学生が履修科目を決定する際の行動に関する研究”, 第 50 回消費者行動研究コンファレンス 報告要旨集, pp. 77-80, 2015.
- [65] Richard E. Petty and John T. Cacioppo, “The Elaboration Likelihood Model of Persuasion”, *Advances in EXPERIMENTAL SOCIAL PSYCHOLOGY*, Vol. 19, pp. 123-205, 1986.
- [66] 清水聰, “新しい消費者行動 (初版第 6 刷)” 千倉書房, 2012.
- [67] 杉本徹雄, “新・消費者理解のための心理学 (初版第 3 刷)” 福村出版, 2013.

- [68] Philip Kotler, 恩藏直人 監修, 月谷真紀 訳, “コトラーのマーケティング・マネジメント 基本編 (初版第5刷)”, ピアソン・エデュケーション, 2004.
- [69] Philip Kotler, 井関利明 (監訳), “非営利組織のマーケティング戦略 - 自治体・大学・病院・公共機関のための新しい変化対応パラダイム- (初版第1刷)”, 第一法規, 1991.
- [70] 清水聰, “戦略的消費者行動論 (初版第2刷)” 千倉書房, 2007.
- [71] Ryotaro Kamimura, Fumihiko Yoshida and Ryozo Kitajima, “Interpreting Cabinet Approval Ratings by Neural Networks, Proceedings of the 25th IASTED International Multi-Conference Artificial Intelligence and Applications (AIA 2007), pp.63-68, 2007.
- [72] Ryotaro Kamimura, Fumihiko Yoshida, Yamashita Toshie and Ryozo Kitajima, “Information-Theoretic Variable Selection in Neural Networks”, Proceedings of the 2007 IEEE Symposium on Foundations of Computational Intelligence (FOCI 2007), pp. 222-227, 2007.
- [73] 北島良三, 上村龍太郎, 内田理, 鳥海不二夫, “ニューラルネットワークを用いた Tweet データの分類に関する研究”, JSAI2015 論文集, pp. 1-4, 2015.
- [74] Ryozo Kitajima, Ryotaro Kamimura, Osamu Uchida and Fujio Toriumi, “Identifying Important Tweets by Considering the Potentiality of Neurons”, IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E99-A, No. 8, pp. 1555-1559, 2016.
- [75] 吉次由美, “東日本大震災に見る大災害時のソーシャルメディアの役割 ～ツイッターを中心に～”, 放送研究と調査, Vol. 7, pp. 16-23, 2011.

- [76] 石川哲也, 川崎昭如, 目黒公郎, “山陰地方豪雪災害時の Twitter ユーザによる情報発信行動に関する分析と考察”, 地域安全学会論文集, Vol. 17, pp. 1-9, 2012.
- [77] 大竹清敬, “対災害 SNS 情報分析システム DISAANA を用いた実証実験”, ITU ジャーナル, Vol. 45, No. 7, pp. 34-37, 2015.
- [78] 水野淳太, 後藤淳, 大竹清敬, 川田拓也, 鳥澤健太郎, クロエツエージュリアン, 田仲正弘, 橋本力, 奥村明俊 “モダリティ解析の導入による対災害 SNS 情報分析システム DISAANA の質問応答性能の改善”, 情報処理学会論文誌 コンシューマ・デバイス & システム, Vol. 6, No. 1, pp. 106-120, 2016.
- [79] 京都大学 大学院情報学研究科 知能情報学専攻 知能メディア講座 言語メディア分野, 黒橋・河原研究室, “日本語形態素解析システム JUMAN version 7.0”, <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>
- [80] 原久美子, 木野泰伸, 鳥海不二夫, “字・町名をキーとした災害時 Twitter 情報の抽出と地図への展開 - 「どこ」で「何」が起きているのかを知る-”, JSAI2013 論文集, pp. 1-4, 2013.
- [81] 榎剛史, 原久美子, 吉田光男, 鳥海不二夫, 篠田孝祐, 栗原聡, 風間一洋, 野田五十樹, “災害情報基盤構築に向けたテキストデータからの地理情報抽出システム”, JSAI2014 論文集, pp. 1-4, 2014.
- [82] 六瀬聡宏, 内田理, 富田誠, 梶田佳孝, 山本義郎, 鳥海不二夫, “大規模災害時の情報提供を目的とした地名の曖昧性解消”, 言語処理学会 第21回年次大会 発表論文集, pp. 222-225, 2015.
- [83] Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier and Sarah Vieweg, “Aidr: Artificial intelligence for disaster response”, Proceedings of the companion publication of the 23rd international conference on World wide web companion, pp. 159-162, 2014.

- [84] Ryoza Kitajima, Ryotaro Kamimura, Osamu Uchida and Fujio Toriumi, “Neural Potential Learning for Tweets Classification and Interpretation”, Proceedings of the 2015 Seventh International Conference of Soft Computing and Pattern Recognition (SoCPaR 2015), pp. 141-148, 2015.
- [85] 山下智志, 三浦翔, “信用リスクモデルの予測精度 -AR 値と評価指標- (初版第 1 刷), 朝倉書店, 2011.
- [86] 北島良三, 上村龍太郎, “社是と企業パフォーマンスの関係”, 第 30 回ファジィ システム シンポジウム講演論文集, pp. 838-843, 2014.
- [87] 北島良三, 上村龍太郎, 内田理, “東証 1 部上場の製造業企業の社是からの特徴抽出に関する研究”, 第 31 回ファジィ システム シンポジウム講演論文集, pp. 306-311, 2015.
- [88] 北島良三, 上村龍太郎, 内田理, “社是に含まれる業績に影響を与える単語に関する研究 -TF-IDF と共起語を切り口にした解析-”, 第 32 回ファジィ システム シンポジウム講演論文集, pp. 463-468, 2016.
- [89] 久保克行, 広田真一, 宮島英昭, “日本企業のコントロールメカニズム : 経営理念の役割”, 季刊 企業と法創造, Vol. 4, pp. 113-124, 2005.
- [90] 松葉博雄, “経営理念の浸透が顧客と従業員の満足へ及ぼす効果 -事例 企業調査研究から-”, 経営行動科学, Vol. 21, No. 2, pp. 89-103, 2008.
- [91] 鳥羽欽一郎, 浅野俊光, “戦後日本の経営理念とその変化 -経営理念調査を手がかりとして-”, 組織科学, Vol. 18, No. 2, pp. 37-51, 1984.
- [92] 並木伸晃, “ミッション・ステートメントと企業業績の関係 -日本化学企業のケース-”, 立教ビジネスレビュー, Vol. 1, pp. 108-113, 2008.
- [93] 小田恵美子, 三橋平, “経営理念と企業パフォーマンス: テキスト・マイニングを用いた実証研究”, 経営哲学, Vol. 7, No. 2, pp. 22-37, 2010.

- [94] 飛田努, “日本企業の組織文化・経営理念と財務業績に関する実証分析 -2000年代における日本的経営を考察する手掛かりとして-”, 立命館経営学, Vol. 48, No. 5, pp. 61-78, 2010.
- [95] Ryotaro Kamimura, Fumihiko Yoshida, Tadanari Taniguti and Ryo-ozo Kitajima, “Inference of Cabinet Approval Ratings by Neural Networks”, Proceedings of SCIS & ISIS 2006, pp. 213-217, 2006.
- [96] 証券コード協議会, “業種別分類項目及び業種コード”, http://www.jpx.co.jp/sicc/category/ct_chart.html
- [97] Ryotaro Kamimura, “Self-Organizing Selective Potentiality Learning to Detect Important Input Neurons”, Proceedings of Systems, Man, and Cybernetics (SMC), pp. 1619-1626, 2015.
- [98] 経済産業省経済産業政策局調査統計部, “我が国の工業 ～変化を続ける製造業～”, <http://www.meti.go.jp/statistics/tyo/kougyo/wagakuni/2011.html>
- [99] Ryo-ozo Kitajima and Ryotaro Kamimura, “Company Policy Analysis By Information Theoretical Neural Networks”, 第40回ファジィ・ワークショップ講演論文集, pp. 13-14, 2014.
- [100] Ryo-ozo Kitajima and Ryotaro Kamimura, “Accumulative Information Enhancement in the Self-Organizing Maps and Its Application to the Analysis of Mission Statements”, Journal of Artificial Intelligence and Soft Computing Research, Vol. 5, No. 3, pp. 161-176, 2015.
- [101] 石田基広, “Rによるテキストマイニング入門 (初版第1刷)”, 森北出版, 2008.
- [102] 内田豪, 佐藤多加之, 北園淳, 岡田真人, 谷藤学, “L1正則化ロジスティック回帰によって明らかにされた下側頭葉視覚連合野における階

層的視覚情報表現”, 電子情報通信学会論文誌 D, Vol. J96-D, No. 7, pp. 1645-1653, 2013.

[103] 農林水産省, “食料品消費モニター調査結果 平成 16 年度特別調査”,
http://www.maff.go.jp/j/heyah/h_monitor/pdf/h1600.pdf

[104] 安土敏, “日本スーパーマーケット原論 (初版第 19 刷)”, ぱるす出版,
2012.

[105] 安土敏, “日本スーパーマーケット創論 内食提供ビジネスのマネジメント (初版第 2 刷)”, 商業界, 2006.

[106] 山本直人, “マーケティング企画技術 (初版第 4 刷)”, 東洋経済新報社,
2008.

[107] 日本マーケティング・リサーチ協会, “第 41 回 (2015 年度) 経営業務
実態調査”, [http://www.jmra-net.or.jp/trend/investigation/
pdf/realities_41/gyoumujitai2016.pdf](http://www.jmra-net.or.jp/trend/investigation/pdf/realities_41/gyoumujitai2016.pdf)

[108] 日本マーケティング・リサーチ協会, “第 36 回 (2010 年度) 経営業務
実態調査”, [http://www.jmra-net.or.jp/trend/investigation/
pdf/realities_36/gyoumujitai2011.pdf](http://www.jmra-net.or.jp/trend/investigation/pdf/realities_36/gyoumujitai2011.pdf)

[109] 日本マーケティング・リサーチ協会, “第 31 回 (2005 年度) 経営業務
実態調査”, [http://www.jmra-net.or.jp/trend/investigation/
pdf/realities_31/gyoumujitai2006.pdf](http://www.jmra-net.or.jp/trend/investigation/pdf/realities_31/gyoumujitai2006.pdf)

業績一覧

※ 2013 年 4 月 1 日からの業績一覧

学術論文

- 1 Ryozo Kitajima, Ryotaro Kamimura, Osamu Uchida and Fujio Toriumi, “Identifying Important Tweets by Considering the Potentiality of Neurons”, IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E99-A, No. 8, pp. 1555-1559, 2016.
- 2 北島良三, 遠藤啓太, 上村龍太郎, “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”, オペレーションズ・リサーチ, Vol. 61, No. 2, pp. 88-96, 2016.
- 3 Ryozo Kitajima, Ryotaro Kamimura, Osamu Uchida and Fujio Toriumi, “Potential Information Maximization: Potentiality-Driven Information Maximization and Its Application to Tweets Classification and Interpretation”, The International Journal of Computer Information Systems and Industrial Management Applications, Vol. 8, pp. 42-51, 2016.
- 4 Ryozo Kitajima and Ryotaro Kamimura, “Accumulative Information Enhancement in the Self-Organizing Maps and Its Application to the Analysis of Mission Statements”, Journal of Artificial Intelligence and Soft Computing Research, Vol. 5, No. 3, pp. 161-176, 2015.

国際会議

- 1 Ryotaro Kamimura, Ryozo Kitajima and Osamu Uchida, “Self-Organized Potential Competitive Learning to Improve Interpretation and Gen-

eralization in Neural Networks”, Proceedings of the Eighth International Conference on Advanced Cognitive Technologies and Applications, pp. 32-37, The Eighth International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2016), 2016.

- 2** Ryoza Kitajima and Ryotaro Kamimura, “Simplifying Potential Learning by Supposing Maximum and Minimum Information for Improved Generalization and Interpretation”, Proceedings of the 35th IASTED International Conference Modelling, Identification and Control (MIC 2016), pp. 20-27, The 35th IASTED International Conference on Modelling, Identification and Control (MIC 2016), 2016.
- 3** Ryotaro Kamimura and Ryoza Kitajima, “SOM Knowledge Induced Learning with Maximum Information Principle to Improve Multi-Layered Neural Networks”, Proceedings of the IASTED International Conference Computational Intelligence (CI 2015), pp. 255-261, The 6th IASTED International Conference on Computational Intelligence (CI 2015), 2015.
- 4** Ryoza Kitajima, Ryotaro Kamimura, Osamu Uchida and Fujio Toriumi, “Neural Potential Learning for Tweets Classification and Interpretation”, Proceedings of the 2015 Seventh International Conference of Soft Computing and Pattern Recognition (SoCPaR 2015), pp. 141-148, The 7th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2015), 2015.
- 5** Ryotaro Kamimura and Ryoza Kitajima, “Selective potentiality maximization for input neuron selection in self-organizing maps”, Proceedings of 2015 International Joint Conference on Neural Networks (IJCNN 2015), pp. 1-8, The International Joint Conference on Neural Networks, 2015.

- 6 Ryozo Kitajima and Ryotaro Kamimura, “Gradual Information Maximization to Clarify the Importance of Company Creeds”, Proceedings of SCIS & ISIS 2014, pp. 1576-1581, Joint 7th International Conference on Soft Computing and Intelligent Systems and 15th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 2014), 2014.
- 7 Ryotaro Kamimura and Ryozo Kitajima, “Gradual Information Maximization in Information Enhancement to Extract Important Input Neurons”, Proceedings of the IASTED International Conference Artificial Intelligence and Applications (AIA 2014), pp. 386-392, The 13th IASTED International Conference on Artificial Intelligence and Applications (AIA 2014), 2014.
- 8 Ryozo Kitajima, Ryotaro Kamimura, “Neural Information Processing Models for Consumer Behaviour Analysis”, Proceedings of MJIT-JUC Joint International Symposium 2013, pp. 1-4, MJIT-JUC Joint International Symposium 2013 (MJJIS2013), 2013.

国内会議

- 1 北島良三, 上村龍太郎, 内田理, “社是に含まれる業績に影響を与える単語に関する研究 -TF-IDF と共起語を切り口にした解析-”, 第 32 回ファジィ システム シンポジウム講演論文集, pp. 463-468, 第 32 回ファジィ システム シンポジウム, 2016.
- 2 北郷翼, 上村龍太郎, 先崎淳一郎, 北島良三, “過剰情報量獲得抑制法の自治体窓口業務改善への応用”, 第 32 回ファジィ システム シンポジウム講演論文集, pp. 381-386, 第 32 回ファジィ システム シンポジウム, 2016.
- 3 先崎淳一郎, 上村龍太郎, 北島良三, 北郷翼, “ニューラルネットワークを用いた板橋区行政サービスの解析”, 第 52 回消費者行動研究コン

ファレンス 報告要旨集, pp. 97-101, 第 52 回消費者行動研究コンファレンス, 2016.

- 4 北島良三, 上村龍太郎, “小売店顧客の継続来店予測モデル作成と継続来店顧客の行動分析 -POS データを入力とするニューラルネットワークを用いた予測モデル-”, Japan Marketing Academy Conference Proceedings, Vol. 4, pp. 266-277, カンファレンス 2015, 2015.
- 5 北島良三, 上村龍太郎, 内田理, “東証 1 部上場の製造業企業の社是からの特徴抽出に関する研究”, 第 31 回ファジィ システム シンポジウム 講演論文集, pp. 306-311, 第 31 回ファジィ システム シンポジウム, 2015.
- 6 北島良三, 上村龍太郎, 内田理, 鳥海不二夫, “ニューラルネットワークを用いた Tweet データの分類に関する研究”, JSAI2015 論文集, pp. 1-4, 2015 年度人工知能学会全国大会 (JSAI2015) , 2015.
- 7 北島良三, 白澤秀剛, 丸山有紀子, 田中真, 坂田圭司, 上村龍太郎, “大学生が履修科目を決定する際の行動に関する研究”, 第 50 回消費者行動研究コンファレンス 報告要旨集, pp. 77-80, 第 50 回消費者行動研究コンファレンス, 2015.
- 8 北島良三, 上村龍太郎, “社是と企業パフォーマンスの関係”, 第 30 回ファジィ システム シンポジウム 講演論文集, pp. 838-843, 第 30 回ファジィ システム シンポジウム, 2014.
- 9 Ryozo Kitajima and Ryotaro Kamimura, “Company Policy Analysis By Information Theoretical Neural Networks”, 第 40 回ファジィ・ワークショップ 講演論文集, pp. 13-14, 第 40 回ファジィ・ワークショップ, 2014.

講演

- 1 ドコモ・イノベーションビレッジ マーケティング勉強会, “AI (ニューラルネットワーク) の POS データへの応用と研究”, 2016.

書籍

- 1 上村龍太郎, 北島良三, 竹内晴彦, 山下俊恵, 吉岡茂, “明日からビジネスで使える! Excel と R によるデータ解析入門”, 丸善出版株式会社, 2014.

賞

- 1 “社是と企業パフォーマンスの関係”, 第 30 回ファジィ システム シンポジウム, 奨励賞, 日本知能情報ファジィ学会, 2015 年 9 月.