

POSデータを用いた牛乳販売量の 需要予測モデルの検討

大同大学情報学部情報システム学科 柏崎優也
指導教員 柴田慎一

目次

- 1.研究背景
- 2.研究目的
- 3.先行研究
- 4.提案手法
- 5.検証方法
- 6.実験概要
- 7.予測結果
- 8.考察
- 9.まとめ
- 10.今後の課題
- 11.参考文献

1. 研究背景 (1/3)

主にコンビニエンスストアやスーパーマーケットでは、

POSシステム (point of sales の略) と呼ばれる管理システムが利用されている。

このPOSシステムから収集できる**POSデータ**は、

商品名や価格、数量、日時などの販売実績情報を収集し、その情報から、

- ・需要予測[1]
- ・購買行動分析[2]

などを行うことができる。



[1] 外園智史[他], "牛乳の企業別需要分析—特売情報を含むPOSデータを利用して—", 2009.

[2] 桑原諭[他], "外的要因を考慮した購買行動分析", 2002.

1. 研究背景 (2/3)

- POSデータにおける需要予測

商品の発注業務は全自動化されていない。
そのため発注者の経験や勘で商品の発注を行うケースも多いため、
必ずしも適切な発注ができているとは言い難い。[3]

破棄によるコストと欠品による購入機会損失は、
トレードオフの関係になっている。
この二つをコントロールして適切な発注量を予測することで
店の利益損失を抑えることができる。



[3]松村直樹[他], "POSデータに基づく欠品時の顧客行動を考慮した小売店舗の購買シミュレーション", 2016.

1. 研究背景 (3/3)

- 需要予測に用いるデータ

需要予測に用いるデータとして牛乳を用いる。

牛乳は日持ちのしない「日配品」であり、

価格変動による販売量の変化が顕著である。

また、日次に出荷され、日次の販売量(明日の販売量)の予測値を得ることは重要である。[4]



[4]高橋幸一[他], ニューラルネットワークモデルによる牛乳販売量予測の検討”, 2000.

2. 研究目的

欠品による購入機会損失



破棄によるコスト

これら2つの損失をコントロールすることで、
店の利益損失を抑えることができる。

そのために、小売店において重要な存在である[3]
毎日発注・納品をする日配品の牛乳を対象として
需要予測を行う。



ニューラルネットワーク、k-NN法、RBFNの3つのモデルを使用して、
需要予測の精度を比較する。

[5]上田隆穂, “牛乳乳製品のマーケティングで3氏が講演”, マーチャンダイジング 세미나

3. 先行研究

- 在庫データ及び売上予測を用いた欠品率の推定に関する研究[6]

売上数量にポアソン分布を想定した対数加法モデル(GAM)の当てはめを用い、商品レベルで欠品率を算出した。

➡ 需要予測モデルの精緻化や、変数選択の検討の必要がある。

- ニューラルネットワークモデルによる牛乳販売量予測の検討[7]

牛乳の販売量に影響を与える要因として、販売価格、曜日、気温が優位なものとなっている。

➡ 時間的制約から変数の選択や時系列モデルなどの検討が行えなかった。

[6]寺嶋正尚[他], "在庫データ及び売上予測を用いた欠品率の推定に関する研究", 2006.

[7]高橋幸一[他], "ニューラルネットワークモデルによる牛乳販売量予測の検討", 2000.

4. 提案手法 (1/3)

●ニューラルネットワーク(神経回路網 neural network)

人間の脳を真似て多数のニューロンを結合したネットワーク上で非線形・並列処理的な情報処理をさせようとする手法。ニューラルネットワークは伝統的な線形回帰分析よりも高い予測性を示すことが多いといわれている。[8]

ニューラルネットワークを用いた研究例

- ・画像認識や音声認識[9]
- ・需要予測

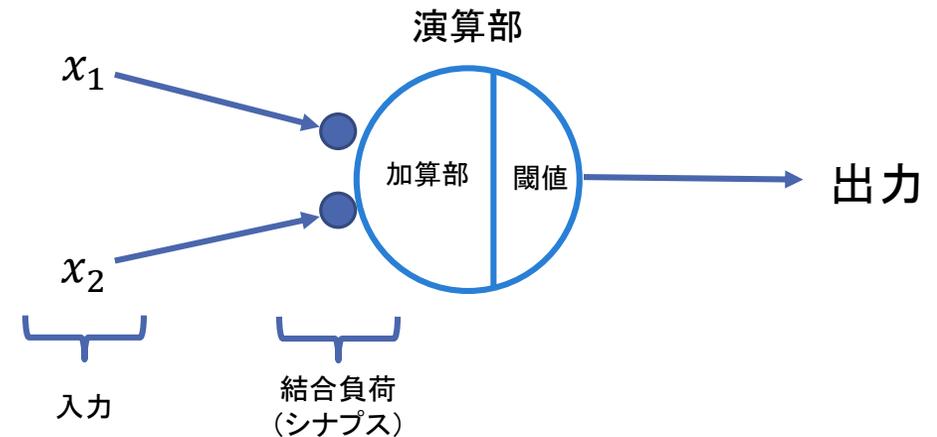


図1 ニューロンモデル

[8]鈴木督久[他], "ニューラルネットによる牛乳販売量の予測", 2001.

[9]綱川菊美, "人工知能ヒストリー: ディープラーニング登場までの軌跡とAI新フェーズ", 2015.

4. 提案手法 (2/3)

●k-NN法 (k-近傍法 k-Nearest Neighbor algorithm)

データ間の距離が近いサンプルをk個取り出し、そのk個の目的変数の値の平均値を推定値として算出する手法である。

データ全体の分布を考慮せずに識別をするため、識別に汎化の影響が少ない。

k-NN法を用いた研究例

- ・頭部動作認識システム[10]
- ・語義曖昧性解消[11]
- ・需要予測

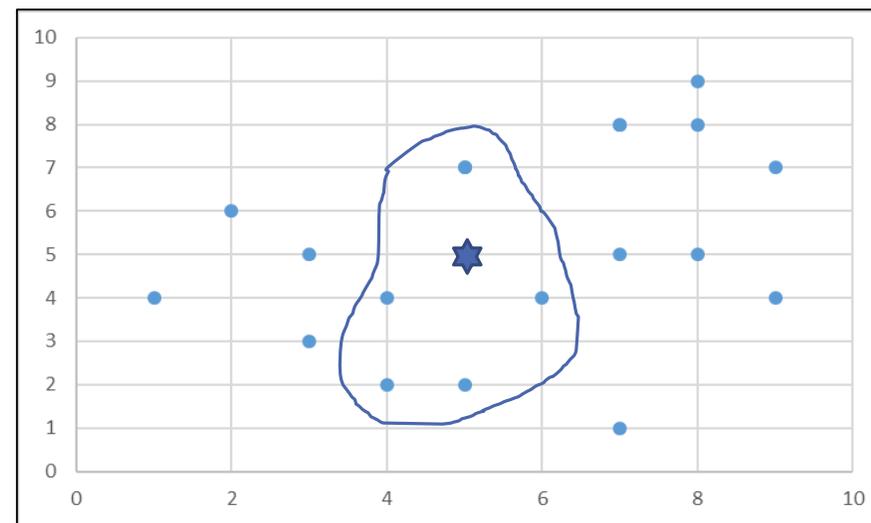


図2.k=5の場合の分類の例

[10]木佐省吾[他], "ウェアラブルセンサを用いた頭部動作認識システムの構築", 2016.

[11]新納浩幸[他], "k近傍法とトピックモデルを利用した語義曖昧性解消の領域適応", 2013.

4. 提案手法 (3/3)

● RBFN (放射基底関数ネットワーク Radial Basis Function Network)

中間層に放射基底関数を用いた3層ニューラルネットワークの一種である。

関数近似能力や学習効率に優れている。[12]

RBFNを用いた研究例

- ・ 顔認識[13]
- ・ 需要予測

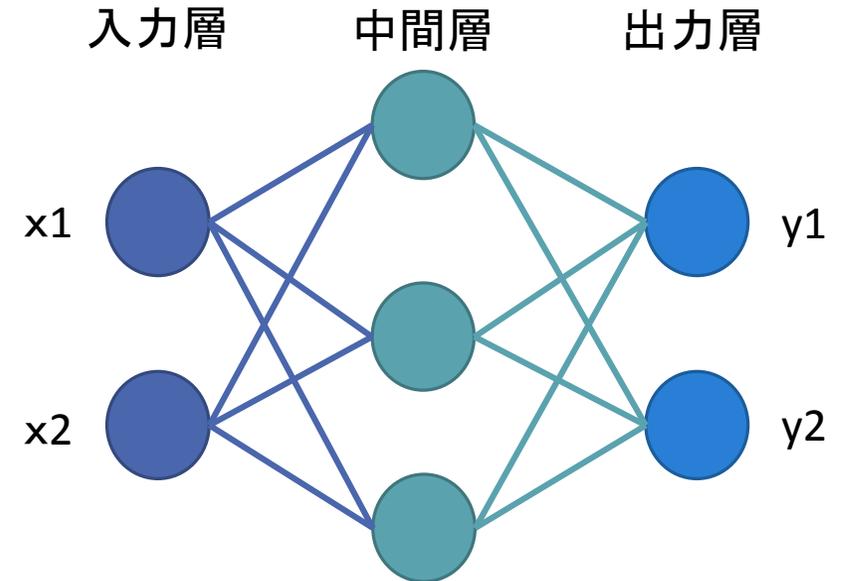


図3. 3層ニューラルネットワークの構造

[12]後藤 一成, "相関を考慮したRBFネットワークを用いたソフトウェアメトリクスデータの分析", 2015.

[13]根岸高弘 [他], "3次元情報を利用したRBFネットワークによる顔認識", 2004.

5. 検証方法

● leave one out 交差検証法

モデルを構築する際の推定誤差を検証する。元のデータを n 個のブロックに分割し、 $n-1$ 個のブロックを 学習データ、学習に使用していない 1 ブロックを 検証用データ としてモデルの精度を算出する。

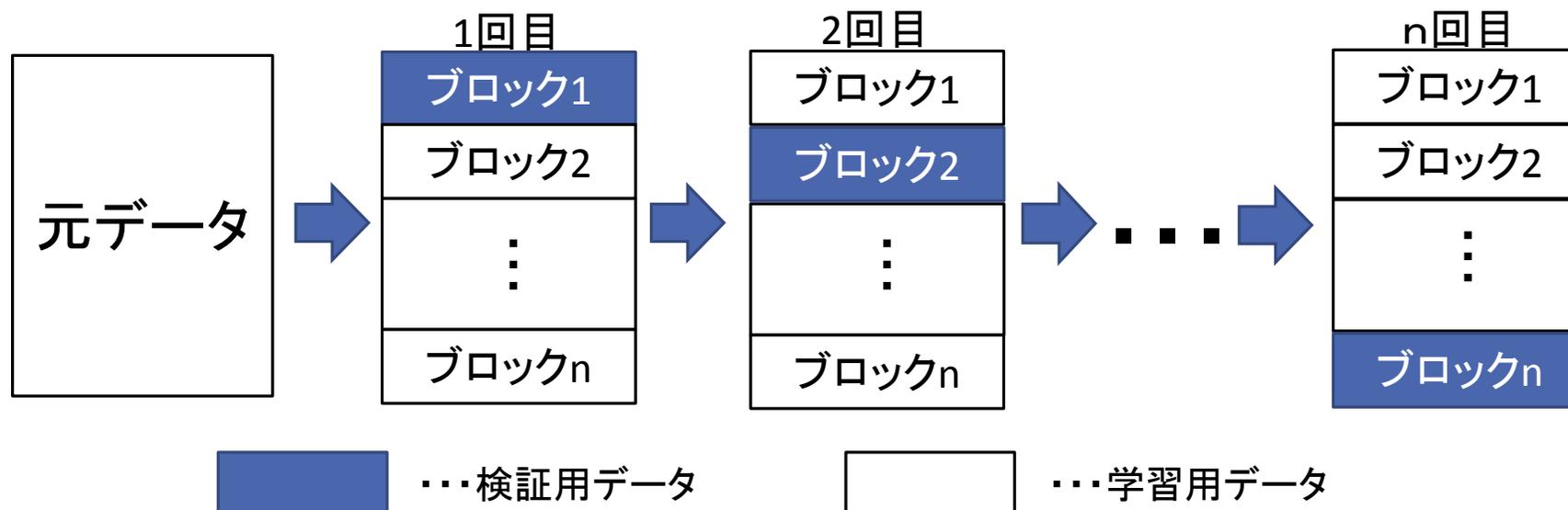


図4. leave one out 交差検証法のモデル

6.使用データ(1/3)

●使用するPOSデータ

実験に使用するデータを以下に記載する。

表1.POSデータ

データ概要	石川県の同一チェーンスーパーマーケット9店舗の販売データ (以下、「店舗1」～「店舗9」と表記)の販売データ
期間	2007年4月～2008年3月までの1年間 (ただし、店舗2の4月のデータが欠損)
データ項目	日付、顧客ID、部門ID、商品名、JANコード、単価の6項目
レコード数	全店舗の平均レコード数は、約45万レコード/月 (販売数の多い店舗では、約75万レコード/月、 販売数の少ない店舗では、約25万レコード/月)

日付	顧客ID	部門ID	商品名	JANコード	単価
2007/4/1 14:50		1	3 紗々ゴールド	4903333039708	198
2007/4/1 14:50		1	3 キシリッシュボト	49771404	282
2007/4/1 14:50		1	3 ミックスゼリー	4901324020216	208
2007/4/1 14:51		2	3 元気になあれ牛乳	4902720048804	168
2007/4/1 14:51		2	7 かぼちゃ	207568000989	98
2007/4/1 14:51		2	2 糸のカミニ3ヶ入	4901160017661	118
2007/4/1 14:51		2	7 雪国もやし	4979754800005	28
2007/4/1 14:51		2	7 さつま芋	5418	148
2007/4/1 14:51		2	9 超熟3枚	4901820332189	89
2007/4/1 14:51		2	7 ニラ	5166	78
2007/4/1 14:52		3	3 明治ラブ1L	4902705080140	168
2007/4/1 14:52		3	7 キャベツ	5159	148
2007/4/1 14:52		3	7 かぼちゃ	207568000989	98
2007/4/1 14:52		3	7 竹の子水煮	4903285003864	198
2007/4/1 14:52		3	2 生芋つきこん20	4971688110557	98

図5. POSデータの最初の部分

6.使用データ(1/3)

- 使用する商品

この1年間販売され続けていた全9店舗の商品から、牛乳の中でも特に売上販売量が高く、互いが競合商品である

「元気になあれ牛乳(以下、[元気牛乳](#)と表記)」

「北海道十勝牛乳(以下、[北海道牛乳](#)と表記)」

を使用する。

6.使用データ(1/3)

全9店舗のうち、
元気牛乳と北海道牛乳が最も販売量が多い「店舗6」と
元気牛乳と北海道牛乳が最も販売量が少ない「店舗8」に
ついて取り組む。

店舗6と店舗8の各商品の販売価格と月平均販売量を下記に示す。

表2.店舗6と店舗8の各商品の販売価格と月平均販売量

	店舗6		店舗8	
	元気牛乳	北海道牛乳	元気牛乳	北海道牛乳
販売価格	148~189円	148~189円	158~189円	158~189円
月平均販売量	約5000個	約4500個	約1000個	約1700個

7. 実験概要 (1/7)

店舗6と店舗8の牛乳の売上個数を予測する

牛乳の中で特に売り上げの高い元気牛乳、北海道牛乳の1日の販売個数の需要予測を行う。

手法としてニューラルネットワーク、k-NN法、RBFNの3つのモデルを使用する。モデルの検証方法として交差検証法を用いる。また、牛乳の販売には定価と特価があるので、各項目の違いや、ヒストグラムでばらつきを見る。

予測結果から、実販売個数と予測個数の相関関係及び誤差から比較して、最良のモデルを調べる。

7. 実験概要 (2/7)

本研究では、2つの実験([実験1](#)、[実験2](#))を行う。

- [元気牛乳](#)の販売個数を予測する(以下、[実験1](#))
- [北海道牛乳](#)の販売個数を予測する(以下、[実験2](#))

実験に用いるモデリング手法が過学習を起こさないよう、説明変数の量を減らすため、[主成分分析](#)を行う。

また、元気個数の販売個数の「[四分位範囲 × 1.7倍](#)」を異常値として除外する。

7. 実験概要 (3/7)

● 各モデル設定

表3.ニューラルネットワーク設定

ニューラルネットワーク設定	
目的関数	Sigmoid-Sum of Square
隠れ層のユニット数	説明変数と同等
繰り返し最大数	1000

表4.k-nn分析 設定

k-NN 設定	
距離関数	Euclid
k値	5

表5.RBFN設定

RBFN 設定	
目的関数	Sigmoid-Sum of Square
係数	1.2
繰り返し最大数	1000

● 交差検証法の設定

- ・ 検証方法.....leave one out交差検証
- ・ データ数.....335個 (店舗6)、324個 (店舗8)

7. 実験概要 (4/7)

●使用する説明変数

実験1、実験2で使用する説明変数を以下に示す。

表6-1.実験1

説明変数	備考	数
曜日	日曜～土曜までを2値で表現	7
平均気温	前日、前々日	2
最高気温	当日、前日、前々日	3
最低気温	当日、前日、前々日	3
降水量	当日、前日、前々日	3
元気牛乳の販売単価	当日、前日、前々日	3
元気牛乳の販売個数	前日、前々日（当日は目的変数）	2
北海道牛乳の販売単価	当日、前日、前々日	3
北海道牛乳の販売個数	前日、前々日	2
説明変数の合計		28

表6-2.実験2

説明変数	備考	数
曜日	日曜～土曜までを2値で表現	7
平均気温	前日、前々日	2
最高気温	当日、前日、前々日	3
最低気温	当日、前日、前々日	3
降水量	当日、前日、前々日	3
元気牛乳の販売単価	当日、前日、前々日	3
元気牛乳の販売個数	前日、前々日	2
北海道牛乳の販売単価	当日、前日、前々日	3
北海道牛乳の販売個数	前日、前々日（当日は目的変数）	2
説明変数の合計		28

7. 実験概要 (5/7)

- 使用するツール

データマイニングツールである“[Visual Mining Studio \(VMS\)](#)”を使用する。

[VMS](#)は株式会社NTTデータ数理システムが開発した、直観的な操作でデータマイニングの処理を行うことのできるツールである。分析前のデータの加工や、多変量解析等の分析機能を数多くそろえている。

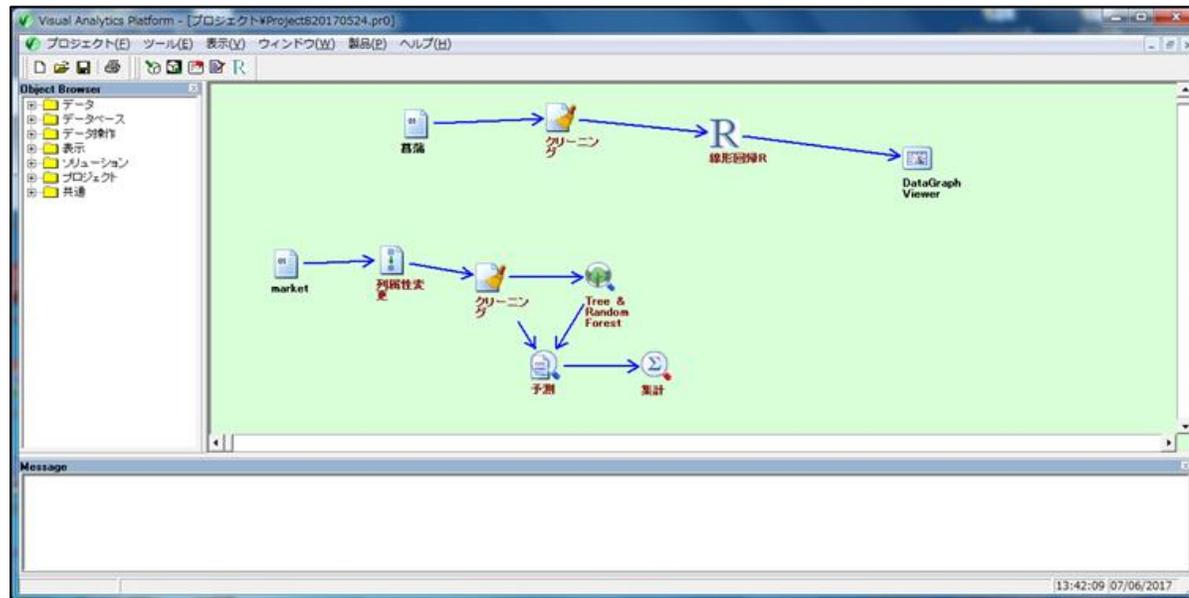


図6 VMSの例



7. 実験概要 (6/7)

- VMSのフロー図

VMSで使用した機能として、下記の機能を使用した。

- ・列属性変更.....データの列の属性を変更する。
 - ・マージ.....複数のデータを結合する。
 - ・主成分分析.....データを主成分で表現する。
 - ・相関.....データ列の相関を求める。
 - ・交差検証法.....モデルを構築する際の推定誤差を検証する。
 - ・データハンドリング.....データの操作、加工を行う。
 - ・データ&グラフビュー.....データをテーブル形式に表示、グラフを作成できる。
- モジュールを使用

これらの機能を一部モジュールを使用して、処理フローをまとめた。
次のページにプロジェクトボードの詳細を記述する。

7. 実験概要 (7/7)

- VMSのフロー図

列属性の変更をして主成分分析をした後に、それぞれのモデルに分けて分析をした。

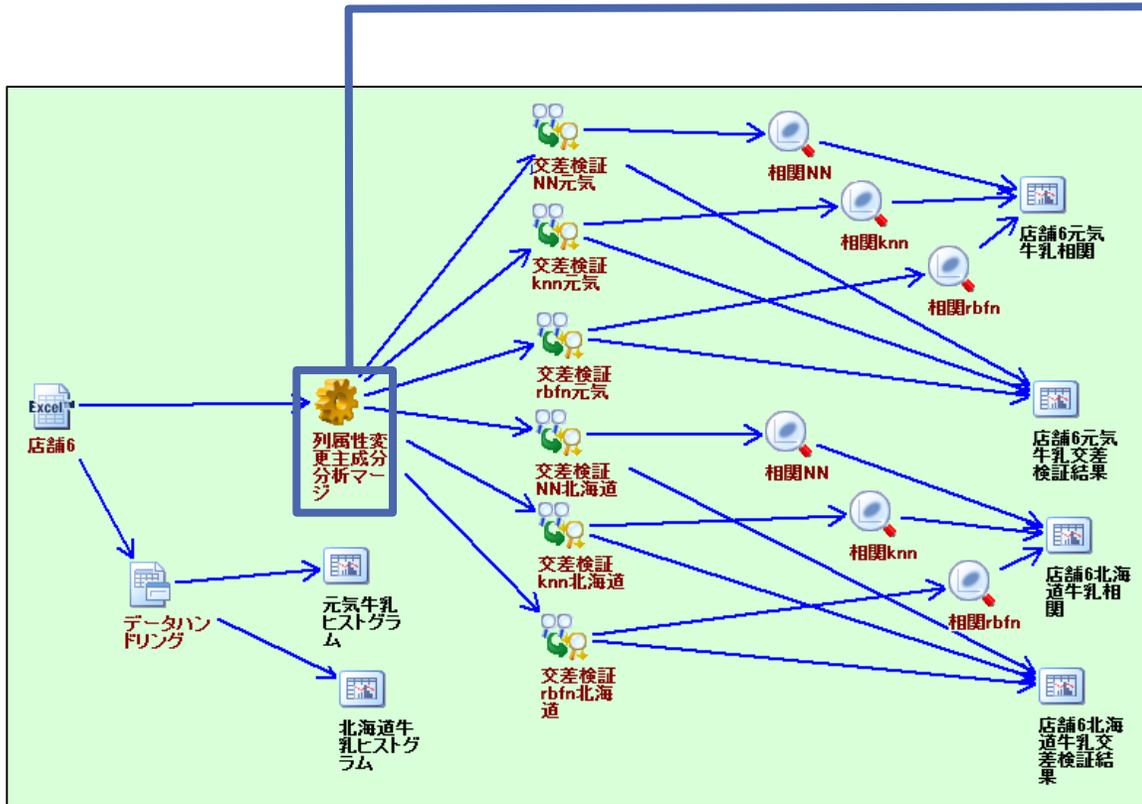


図7.フロー図の一部

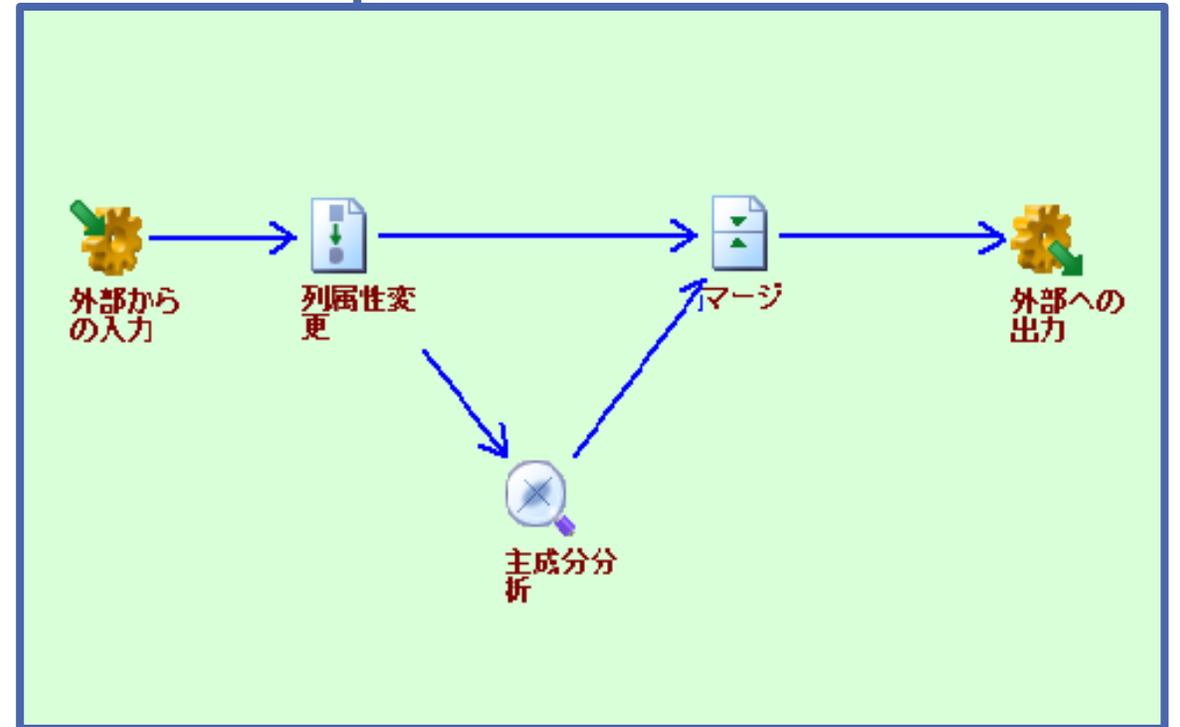


図8.モジュール内のフロー図

7.実験1結果(1/8)

表9.実験1予測結果

		ニューラルネットワーク	k-NN法	RBFN
相関係数		0.964	0.957	0.926
誤差	平均	22	24	31
	最大値	106	104	222
	標準偏差	19	20	29
累計誤差		7349	8098	10327

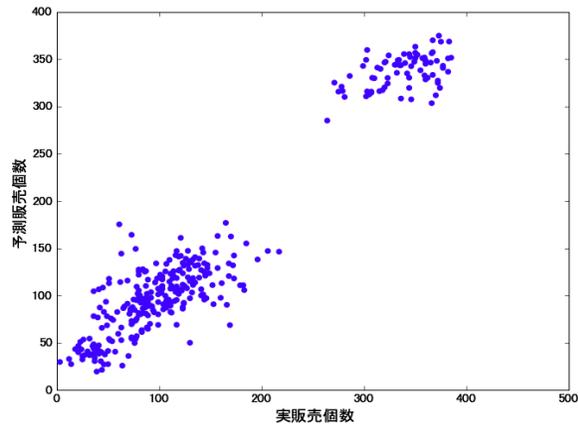


図9-1 ニューラルネットワーク散布図

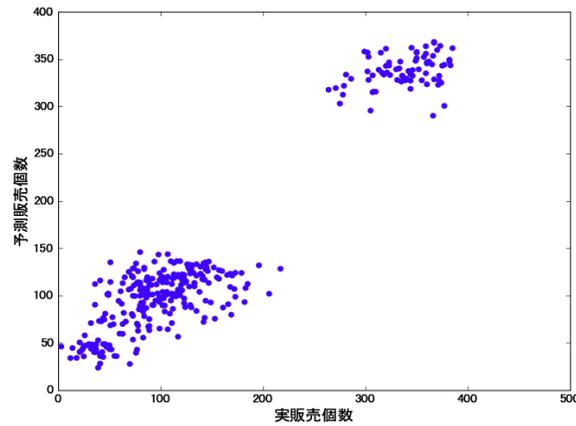


図9-2 k-NN法散布図

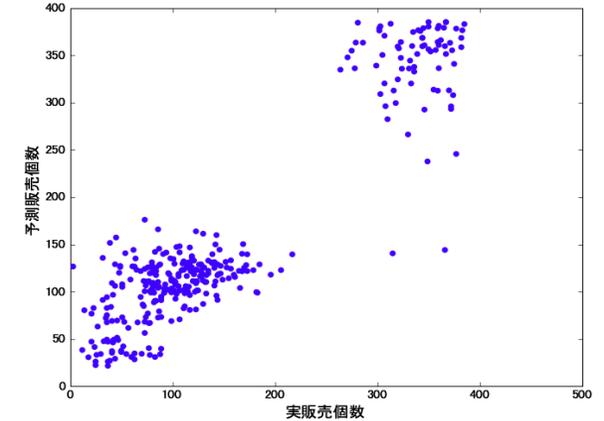


図9-3 RBFN散布図

元気牛乳はニューラルネットワークが、相関係数は最も高く、誤差の平均や標準偏差、累計誤差が最も小さい結果となった。

7. 実験1結果 (2/8)

実験1の定価と特価の表とヒストグラムを下記に記載する。
なお、ヒストグラムは上が定価、下が特価である。

表10. 実験1の定価と特価の表

	定価	特価
販売日数 (日)	93	242
値段 (円)	~189	~168
販売平均本数 (本)	64	184
販売最大本数 (本)	130	385
標準偏差	30	110
月販売量の比	0.117	0.883

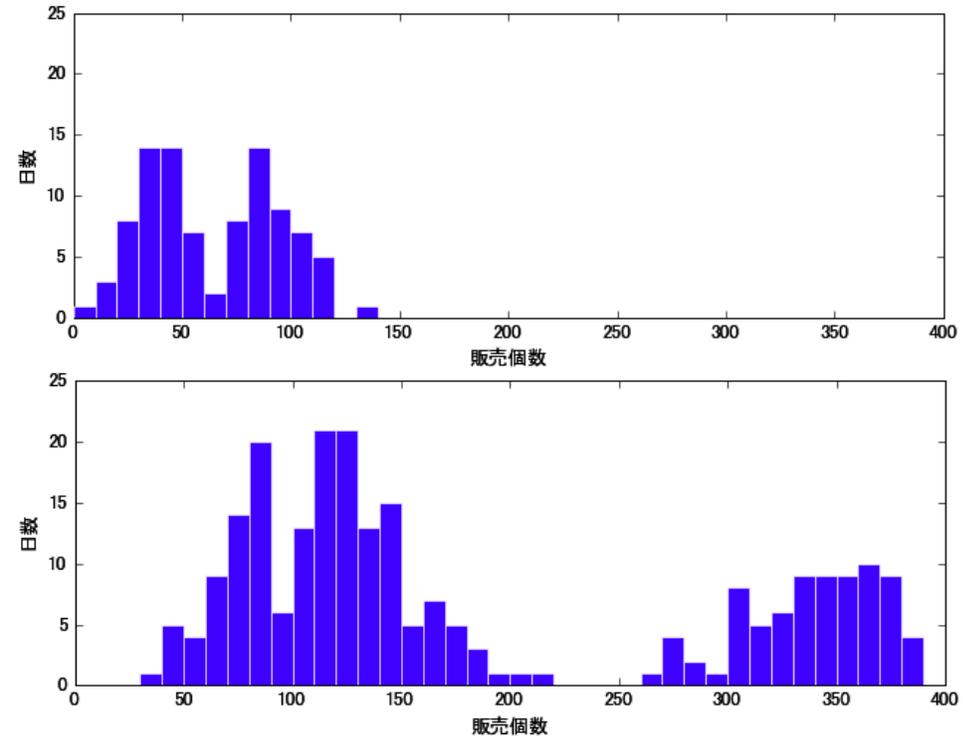


図10. 実験1の定価と特価のヒストグラム

7.実験1結果(3/8)

店舗8

表11.実験1予測結果

		ニューラルネットワーク	k-NN法	RBFN
相関係数		0.897	0.875	0.839
誤差	平均	12	13	15
	最大値	67	58	59
	標準偏差	10	11	12
累計誤差		3936	4319	4942

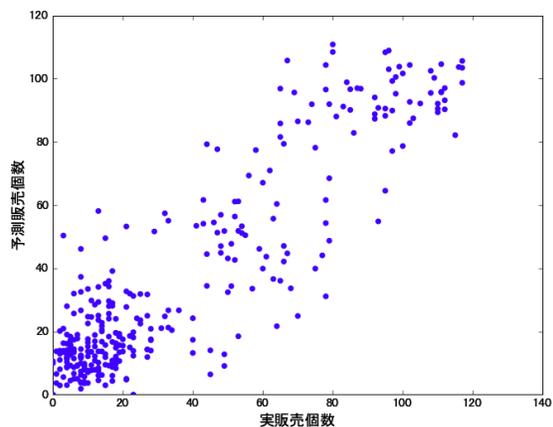


図11-1 実験1ニューラルネットワーク散布図

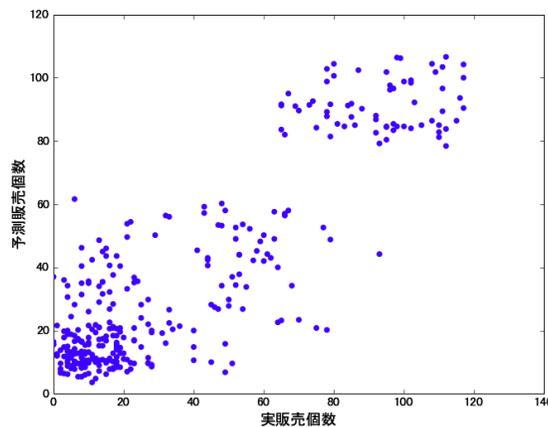


図11-2 実験1k-NN法散布図

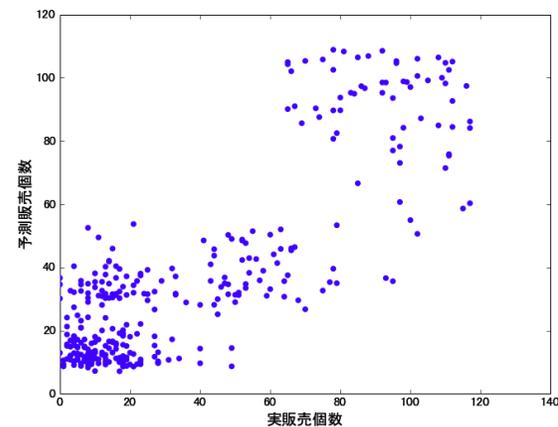


図11-3 実験1RBFN散布図

北海道牛乳はニューラルネットワークが相関係数は最も高く、誤差の平均や標準偏差、累計誤差が最も小さい結果となった。

7. 実験1結果 (4/8)

実験1の定価と特価の表とヒストグラムを下記に記載する。
なお、ヒストグラムは上が定価、下が特価である。

表12. 実験1の定価と特価の表

	定価	特価
販売日数 (日)	210	125
値段 (円)	~189	~168
販売平均本数 (本)	82	273
販売最大本数 (本)	200	480
標準偏差	42	92
月販売量の比	0.354	0.646

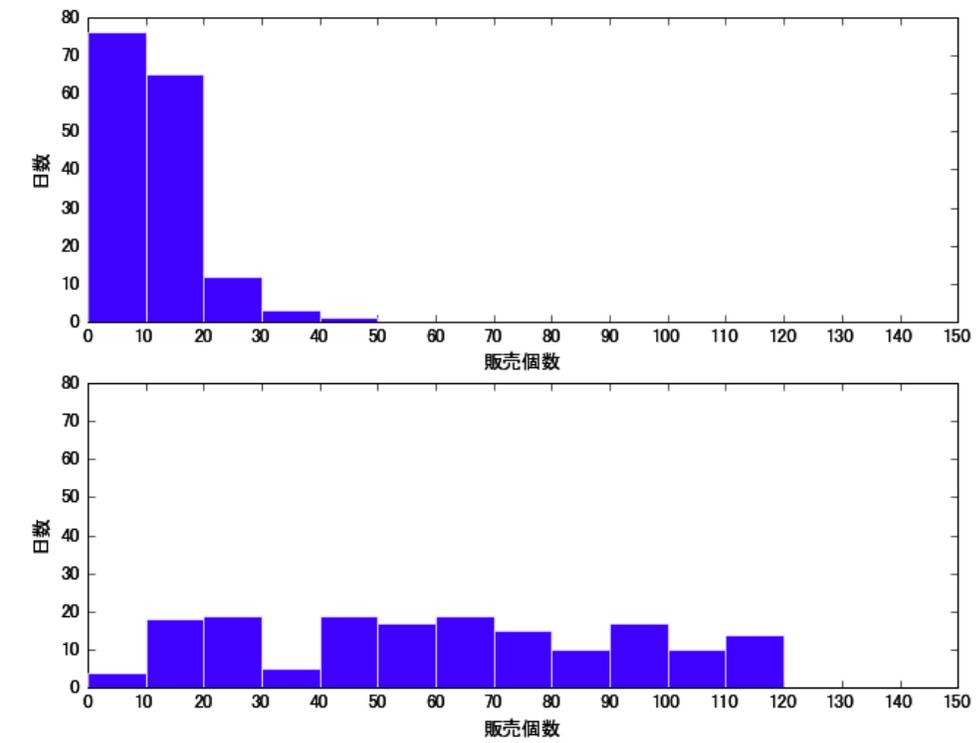


図12. 実験1の定価と特価のヒストグラム

7.実験2結果(5/8)

店舗6

表13.実験2予測結果

		ニューラルネットワーク	k-NN法	RBFN
相関係数		0.949	0.946	0.882
誤差	平均	25	26	39
	最大値	147	206	237
	標準偏差	26	27	37
累計誤差		8366	8558	13130

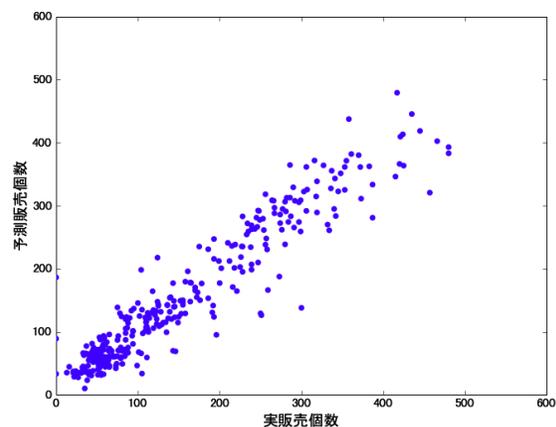


図13-1 実験2ニューラルネットワーク散布図

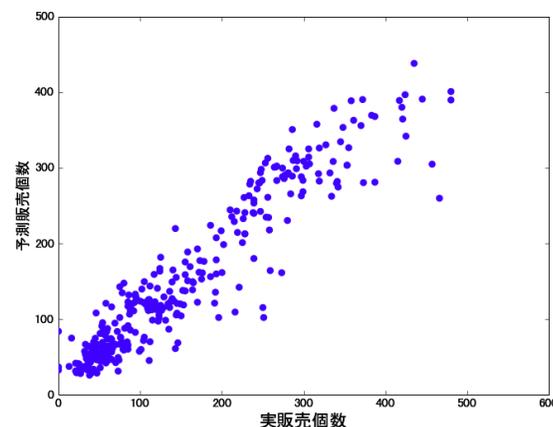


図13-2 実験2k-NN法散布図

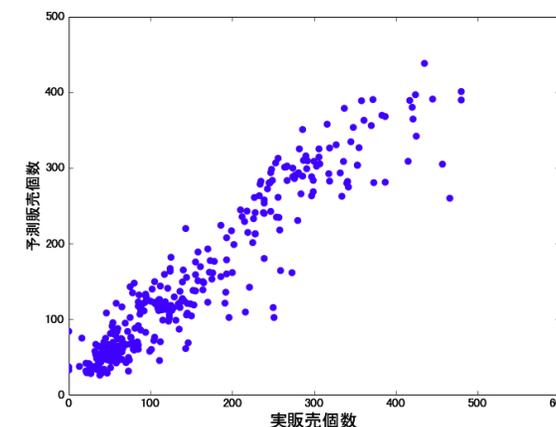


図13-3 実験2RBFN散布図

元気牛乳は、ニューラルネットワークが相関係数は最も高く、誤差の平均や標準偏差、累計誤差が最も小さい結果となった。

7. 実験2結果 (6/8)

実験2の定価と特価の表とヒストグラムを下記に記載する。
なお、ヒストグラムは上が定価、下が特価である。

表14. 実験2の定価と特価の表

	定価	特価
販売日数 (日)	157	167
値段 (円)	~189	~168
販売平均本数 (本)	11	61
販売最大本数 (本)	40	117
標準偏差	7	32
月販売量の比	0.147	0.853

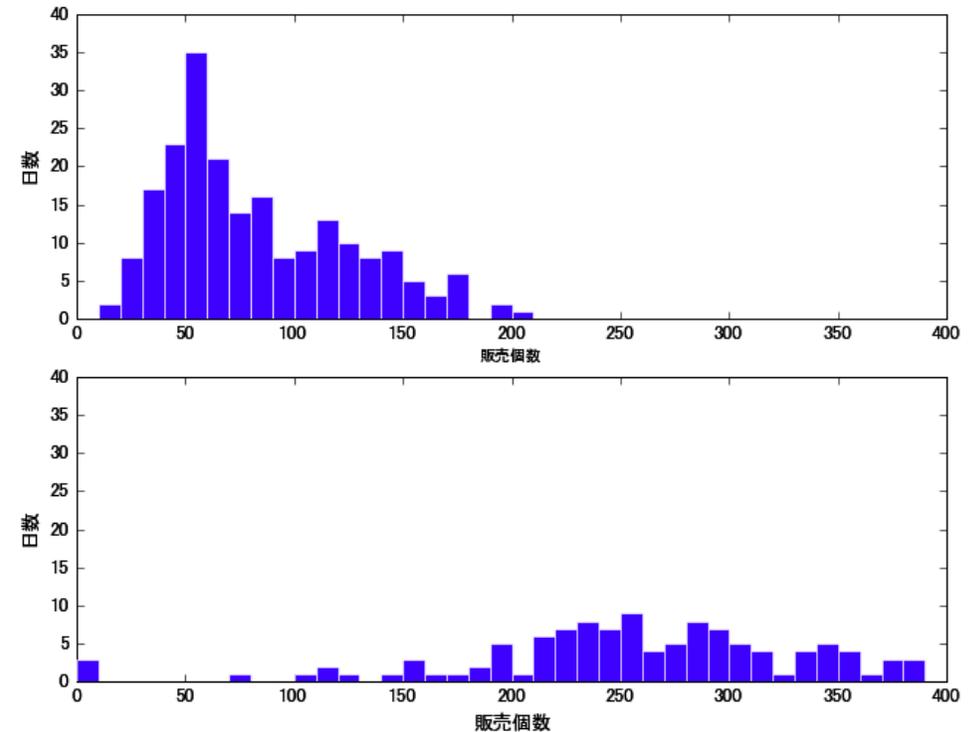


図14. 実験2の定価と特価のヒストグラム

7.実験2結果(7/8)

店舗8

表15.実験2予測結果

		ニューラルネットワーク	k-NN法	RBFN
相関係数		0.940	0.939	0.920
誤差	平均	11	11	15
	最大値	80	84	59
	標準偏差	11	11	11
累計誤差		3776	3794	5018

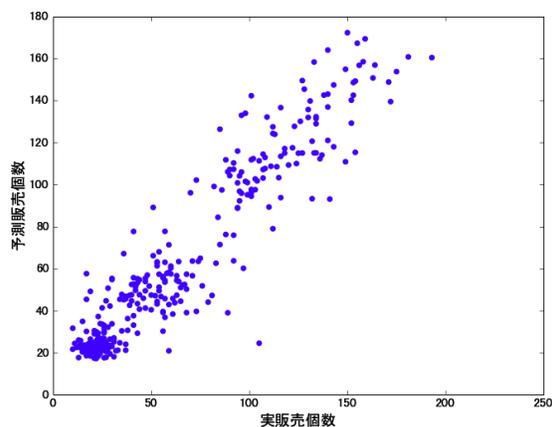


図15-1 実験2ニューラルネットワーク散布図

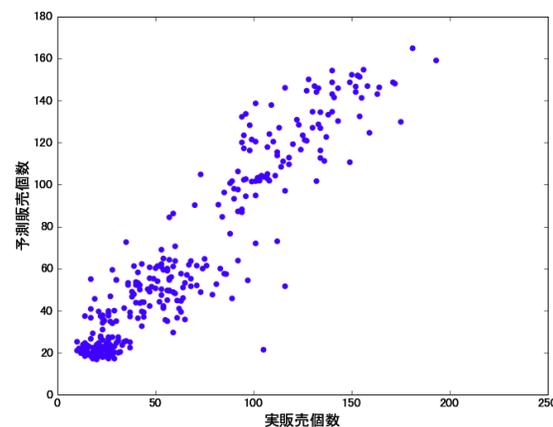


図15-2 実験2k-NN法散布図

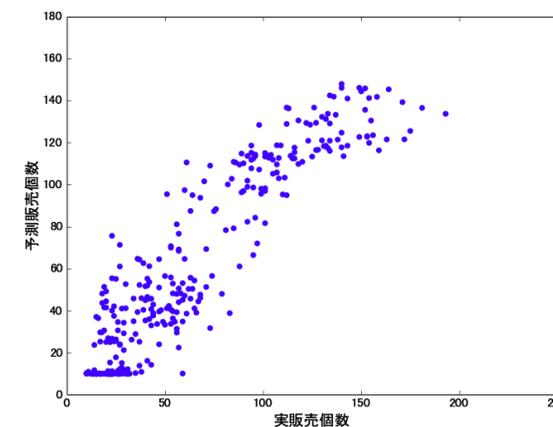


図15-3 実験2RBFN散布図

北海道牛乳はニューラルネットワークが、相関係数は最も高く、誤差の平均や標準偏差、累計誤差が最も小さい結果となった。

7. 実験2結果 (8/8)

実験1の定価と特価の表とヒストグラムを下記に記載する。
なお、ヒストグラムは上が定価、下が特価である。

表16. 実験2の定価と特価の表

	定価	特価
販売日数 (日)	208	116
値段 (円)	~189	~168
販売平均本数 (本)	36	117
販売最大本数 (本)	83	193
標準偏差	18	28
月販売量の比	0.354	0.646

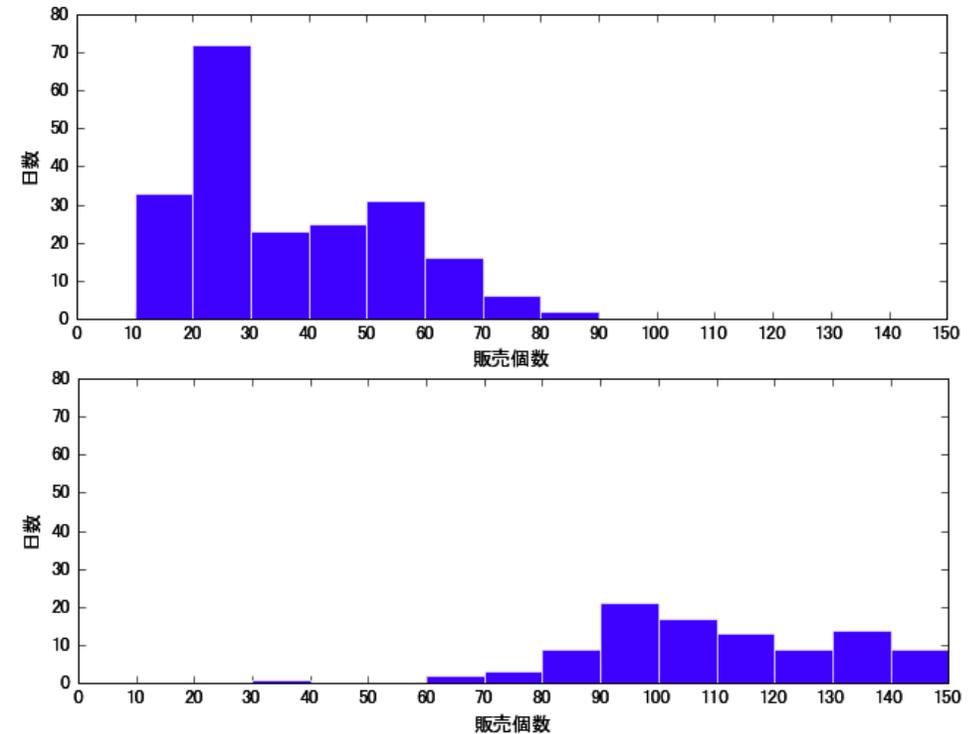


図16. 実験2の定価と特価のヒストグラム

8. 考察 (1/2)

実験1では、相関係数を見るとニューラルネットワークが相関係数が最も高く、累計誤差も最も低い。

また、実験2でもニューラルネットワークが相関係数が最も高く、累計誤差も最も低い。これより、本実験の需要予測のモデリング手法としてニューラルネットワークが最良のモデリング手法であるといえる。

これは、POSデータがニューラルネットワークが学習を行うのに十分なデータがあったと思われる。

しかし、RBFNの結果が悪いため今後はモデリング手法の設定についても考えなければならない。

8. 考察 (2/2)

定価と特価について比べた結果、
元気牛乳の特価日は両店舗ともに8割、北海道牛乳は6割と
両商品ともに特価日が定価日を上回っている。さらに特価日は商品の販売平均量も多い。
これにより、牛乳の売れ行きは特価が重要であるといえる。

実験1の店舗6については定価の販売個数と同じ本数程度しか売れていない
特価日は147日あり特価日全体の約6割である。

しかし実験2の店舗6は21日あり特価日全体の約1割である。

これは店舗6では元気牛乳と北海道牛乳は競合商品ではあるが、北海道牛乳は
特価日のみ買う客が多いためであるといえる。

しかし、店舗8は元気牛乳が特売日のみに買う客が多い。

このため、店舗によって牛乳の販売量に特色があるといえる。

9.まとめ

破棄によるコストと欠品による購入機会損失をコントロールして適切な発注量を予測することで店の利益損失を抑えることができる。そのため、POSデータを用いて需要予測を行う。

需要予測にニューラルネットワーク、k-NN法、RBFNを用いて販売個数の予測を行い、予測精度を検証した。

結果として、実験1、実験2ともにニューラルネットワークが相関係数が最も高く、累計誤差も最も低い。

これより、本実験の需要予測のモデリング手法としてニューラルネットワークが最良のモデリング手法であるといえる。

また、定価と特価の違いにおいて、店舗によってばらつきがあるが、牛乳の売上にとって特価は重要な存在になることがわかった。

9. 今後の課題

- 牛乳の販売シェアの約7割は特売で占められている[2]。
そのため牛乳の販売にとって特売は重要な役割があると思われるので、
特売についての研究もより多く取り入れていきたい。
- 各モデルの予測傾向が他店舗でも同じかどうかを検討する。
- 説明変数の削減方法を主成分分析だけでなく、
決定木なども取り入れて実験を行う。
- 各モデルの設定や説明変数の選択について吟味し、
予測精度の向上を目指す。

10.参考文献

- [1]外園智史, 狩野秀之, 前田幸嗣, “牛乳の企業別需要分析—特売情報を含むPOSデータを利用して—”, フードシステム研究, Vol.16, No.3, pp.15-23, 2009.
- [2]桑原諭, 椿広計, “外的要因を考慮した購買行動分析”, 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集, pp.162-163, 2002.
- [3]松村直樹, 和泉潔, 山田 健太, “POSデータに基づく欠品時の顧客行動を考慮した小売店舗の購買シミュレーション”, 人工知能学会論文誌, Vol.31, No.2, F-F13, pp.1-8, 2016.
- [4]高橋幸一, 石川征郎, “ニューラルネットワークモデルによる牛乳販売量予測の検討”, 日本SASユーザー会総会および研究発表会論文集, Vol.19, pp.235-242, 2000.
- [5]上田隆穂, “牛乳乳製品のマーケティングで3氏が講演”, マーチャンダイジング 세미나, 酪農情報, pp.4-7, 2008.
- [6]寺嶋正尚, 生田目崇, “在庫データ及び売上予測を用いた欠品率の推定に関する研究”, 日本物流会誌, Vol.14, pp.93-100, 2006.
- [7]伊藤 房雄, “牛乳・乳製品の需要分析”, 農業研究報告書, Vol.26, pp.153-162, 1993.
- [8]鈴木督久, “ニューラルネットによる牛乳販売量の予測”, 品質管理, Vol.52, No.3, pp.249-254, 2001.
- [9]綱川菊美, “人口知能ヒストリー: ディープラーニング登場までの軌跡とAI新フェーズ”, 福祉社会学部論文集, Vol.34, No.2, pp.65-74, 2015.
- [10]木佐省吾, 堀内 匡, “ウェアラブルセンサを用いた頭部動作認識システムの構築”, 知能と情報, Vol.28, No.6, pp.986-991, 2016.
- [11]新納浩幸, 佐々木稔 “k近傍法とトピックモデルを利用した語義曖昧性解消の領域適応”, 自然言語処理, Vol.20, No.5, pp.707-726, 2013.
- [12]後藤 一成, “相関を考慮したRBFネットワークを用いたソフトウェアメトリクスデータの分析”, 法政大学大学院紀要 理工学・工学研究科編, Vol.56, 2015.
- [13]根岸高弘, 山下哲孝, 呂建明, 関屋大雄, 谷萩隆嗣, “3次元情報を利用したRBFネットワークによる顔認識”, 2004年電子情報通信学会基礎・境界ソサイエティ大会, pp.136, 2004.

ご清聴ありがとうございました