

ニューラルネットワークによる 体感温度を用いた牛乳販売量予測

大同大学 情報学部 情報システム学科

松田紗季

目次

- p.03 --- 1.研究背景
- p.06 --- 2.関連研究
- p.08 --- 3.提案手法
- p.12 --- 4.使用データ
- p.12 --- 5.実験概要
- p.19 --- 6.実験結果
- p.21 --- 7.考察・まとめ
- p.25 --- 8.今後の課題
- p.26 --- 9.参考文献

1. 研究背景(1/3)

商品の発注, 在庫管理について.

商品の発注は, 全自動化されておらず,
発注者の経験や勘により発注を行うケースが多い. [1]
発注者の負担は大きく, 適切な発注業務が行われているとは限らない.

適切な発注でないとき…,

廃棄による損失 ← **トレードオフ** → 欠品による購入機会の損失

商品の発注量を増やす…欠品は減少するが, 廃棄による損失は増加.
商品の発注量を減らす…廃棄は減少するが, 欠品による機会損失は増加.

[1]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(2/3)

POSシステム (Point Of Sales system: 販売時点情報管理)

コンビニエンスストアやスーパーマーケットなどの小売店で用いられる、商品の販売情報の管理システム。

POSデータのもつ、商品名や金額などの商品の情報や、数量、売れた日時、販売地域などの情報から販売戦略を立てることができる。

いつ、どこで、何が、
どれだけ売れた？



1. 研究背景 (3/3)

POSデータを用いて、牛乳の販売数量を予測する研究.

牛乳は、「日配品」とよばれる、あまり日持ちがしない
要冷蔵商品であり、その中でも購入頻度が高い[2]と考えられる。
日配品は、新鮮なうちに売り切ることが望まれる。[3]

POSデータを用いて、販売数量を予測することで、
余剰在庫による廃棄の損失、
欠品による販売機会の損失の問題を解消する。



[2]福島和宏[他], "世界金融危機による消費者の食品購買行動の変化--日別POSデータによる分析", 2011.

[3]鈴木督久[他], "POSデータのデータマイニング", 2000.

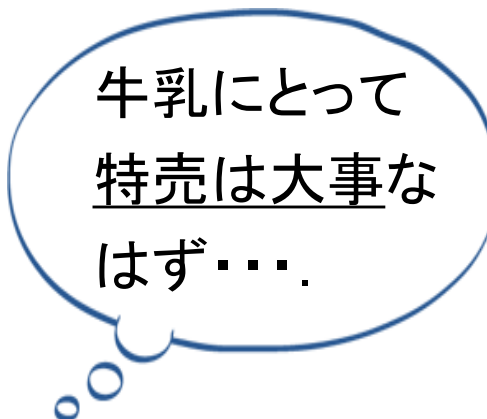
2. 関連研究(1/2)

ニューラルネットによる牛乳販売量の予測[4]

特売日や大晦日などの人為的な特異日は、
予測モデル構成にとっては除外して問題ないと判断している。

牛乳の企業別需要分析[5]では・・・

分析対象データの販売金額シェアの約7割が特売。
牛乳は品質格差が少なく、差別化戦略がとりにくい。
牛乳は特売などの価格競争に陥りがち。



牛乳にとって
特売は大事な
はず・・・

(乳業の構造[6](1991年)にて差別化に成功しつつあるといわれるが、2009年時点では、
まだ、業界第1位のメーカーでさえ差別化が十分とは言えない結果となっている。)

[4]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.

[5]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.

[6]矢坂 雅充, "乳業の構造", 1991.

2. 関連研究(2/2)

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]

POSデータ, 気象データ, 経済統計データ(有効求人倍率・所定労働時間)等を説明変数に用いるが,
時間的制約により変数選択を十分に行えなかった。

POSデータのデータマイニング[8]

牛乳を分析対象とし, ニューラルネットワークを用いている。
・・・しかし, 「分析結果は, 当日プレゼンテーションする。」とあり,
肝心な「分析結果」を確認することができなかった。

[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.

[8]鈴木 督久[他], "POSデータのデータマイニング", 2000.

3. 提案手法(1/4)

ニューラルネットワーク(NN=Neural Network: 神経回路網)による予測.

人間の脳の神経回路の仕組みを模したモデル.

コンピュータに学習能力を持たせ, 人間の脳のメカニズムを人工的に実現することで, 様々な問題を解決する. [9]

ニューラルネットワークを使用した研究例

- ・サクラの開花日の予測[10]
- ・文字認識[11]



[9] "ニューラルネットワーク", <http://www.geocities.co.jp/SiliconValley-Cupertino/3384/nn/NN.html>.

[10] 高野 哲夫, "ニューラルネットワークによるサクラ開花日の学習・予測実験: 山形・新潟を例に(調査ノート)", 2015.

[11] 山村 豊[他], "ニューラルネットワークによる文字認識", 2013.

3.提案手法(2/4)

ニューラルネットワーク(NN=Neural Network: 神経回路網)による予測.

伝統的な線形重回帰分析よりも高い予測成績を示すことが多いといわれる. [4]

データマイニング用ソフトウェアに標準的に装備され,
容易に利用できるツールである. [4]

昨年度に本研究室で行われた,
ニューラルネットワーク・k-NN法・RBFネットワークの
3つのモデルを比較した実験でも,
ニューラルネットワークの予測精度が最も高かった.

3.提案手法(3/4)

使用するツール: VMS (Visual Mining Studio)

VMSは、データマイニングに必要な様々な処理をグラフィカルに行うことのできる統合解析ツールである。

VMSは、データ解析・マイニング・機能に特化した VAPのパッケージである。

VAP (Visual Analytics Platform) は、NTTデータ数理システムが持つ強力なデータマイニング・数理計画・金融等、様々な意思決定問題に対するソリューションを統合的に扱う目的で開発された製品である。

3.提案手法(4/4)

体感温度

売上予測に体感温度を用いている研究[12]があった。

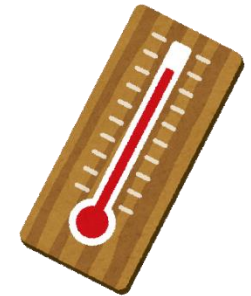
体感温度は客の心理に大きな影響を与える[13]とも言われている。

風速による体感温度 → リンケの体感温度 = $t - 4\sqrt{V}$

湿度による体感温度 → ミスナールの体感温度 = $t - \frac{1}{2.3}(t - 10)(0.8 - \frac{h}{100})$

t = 気温[°C]
V = 風速[m/s]
h = 湿度[%]

気温, リンケの体感温度, ミスナールの体感温度 をそれぞれ説明変数に用いて予測し, 結果を比較する。



[12]竹村 遼[他], "外食産業の売上予測について - 外食チェーンのある店舗における分析結果より - ", 2012.

[13]緒方 知之(編著), "鈴木敏文 考える経営", 2005.

4.使用データ (1/3)

使っているPOSデータ

石川県の同一チェーンスーパーマーケット9店舗の、1年間 (2007年4月～2008年3月)の月別データ。

	A	B	C	D	E	F
1	日付	顧客ID	部門ID	商品名	JANコード	単価
2	2007/4/1 14:50	1	3	紗々ゴールド	4903333039708	198
3	2007/4/1 14:50	1	3	キシリッシュボト	49771404	282
4	2007/4/1 14:50	1	3	ミックスゼリー	4901324020216	208
5	2007/4/1 14:51	2	3	元気になあれ牛乳	4902720048804	168
6	2007/4/1 14:51	2	7	かぼちゃ	207568000989	98
7	⋮			⋮		

ひと月で、
数十万件。

図1 Excelで開いたPOSデータ

4.使用データ (2/3)

この中から、全9店舗で1年間販売され続けていた商品のうち、
 販売量上位3商品の「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)」、
 「農協牛乳」について調査中。

表1 販売量上位3商品

	元気牛乳	北海道牛乳	農協牛乳
定価の価格	~189円	~189円	208円
特価の価格	~168円	~168円	188円
特価日(曜日)	土・日	火・木	毎月第1金
月平均販売量	約2500個	約3100個	約630個

特価日はどちらも
週に2日あり、
価格帯も似ている。

この中でも特に販売量が多いのが、元気牛乳と北海道牛乳。
 農協牛乳とは差が大きく、この2商品は競合商品であり、主力商品。

4.使用データ (3/3)

店舗の特徴

全9店舗(以下,「店舗1」~「店舗9」と表記)のうち牛乳の販売量に特徴がある店舗は…,

店舗1 : 元気牛乳も北海道牛乳も, 平均的な販売量である.

店舗6 : 元気牛乳も北海道牛乳も, 他店舗より販売量が多い.

店舗8 : 元気牛乳も北海道牛乳も, 他店舗より販売量が少ない.

店舗9 : 元気牛乳より, 北海道牛乳が2倍ほど販売量が多い.
また, 常に特価の価格で販売されている.

データの量が多いので, まずは販売量が多い商品,
特徴がある店舗について取り組んでいる.

5. 実験概要 (1/4)

ニューラルネットワークを利用して販売量を予測.

実験概要 (対象: 店舗6の1年間の北海道牛乳/元気牛乳)

互いに競合商品である 北海道牛乳 と 元気牛乳 のある1日の販売個数を

① 気温, ② リンケの体感温度, ③ ミスナールの体感温度 をそれぞれ説明変数に用いて予測する.

予測結果より, 実販売個数と予測個数の相関係数と, 実販売個数と予測個数の差(誤差)を比較する.

前日の情報は予測に影響を与えているか?

また, 変数に前日・前々日の変数があるものについて, 以下の条件で比較する.

- ・ 前日と前々日の変数を用いる場合(=「前日あり」).
- ・ 前日と前々日の変数を用いない場合(=「前日なし」).

5. 実験概要 (2/4)

ニューラルネットワークを利用して販売量を予測.

表2 設定項目

交差検証	
アルゴリズム	Neural Network
交差検証方法	Leave one out
学習	
目的関数	Sigmoid-Sum of Square
隠れ層のユニット数	説明変数と同数 気温:26(前日なし:10) 体感温度:20(前日なし:8)
繰り返し最大数	1000
目的変数	北海道牛乳, または 元気牛乳の“販売個数”

説明変数は, 次ページ以降.

ニューラルネットワークの設定は,
今回は, 表2のように指定した.

(今後, 設定の変更や,
新たな説明変数の追加を行うことで,
予測精度の向上を目指す.)

5.実験概要(3/4)

表3 気温の説明変数(北海道)

説明変数	備考	数
曜日		1 (1)
元気牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
元気牛乳の販売個数	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の販売個数	前日, 前々日の2つ(当日は目的変数).	2 (1)
来客数	前日, 前々日の2つ(当日は未知).	2 (1)
平均気温	当日, 前日, 前々日の3つ.	3 (1)
最高気温	当日, 前日, 前々日の3つ.	3 (1)
最低気温	当日, 前日, 前々日の3つ.	3 (1)
降水量	当日, 前日, 前々日の3つ.	3 (1)
		計:26(10)

表4 気温の説明変数(元気)

説明変数	備考	数
曜日		1 (1)
元気牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
元気牛乳の販売個数	前日, 前々日の2つ(当日は目的変数).	2 (1)
北海道牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の販売個数	当日, 前日, 前々日の3つ.	3 (1)
来客数	前日, 前々日の2つ(当日は未知).	2 (1)
平均気温	当日, 前日, 前々日の3つ.	3 (1)
最高気温	当日, 前日, 前々日の3つ.	3 (1)
最低気温	当日, 前日, 前々日の3つ.	3 (1)
降水量	当日, 前日, 前々日の3つ.	3 (1)
		計:26(10)

※数の()の中の数字は前日なしの場合.

※当日の変数がないものは, 前日の値を使用し, 前々日の値を使用しない.

5. 実験概要(4/4)

表5 リンケの体感温度の説明変数(北海道)

説明変数	備考	数
曜日		1 (1)
元気牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
元気牛乳の販売個数	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の販売個数	前日, 前々日の2つ(当日は目的変数).	2 (1)
来客数	前日, 前々日の2つ(当日は未知).	2 (1)
リンケの体感温度	当日, 前日, 前々日の3つ.	3 (1)
降水量	当日, 前日, 前々日の3つ.	3 (1)
		計: 20(8)

表6 ミスナールの体感温度の説明変数(北海道)

説明変数	備考	数
曜日		1 (1)
元気牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
元気牛乳の販売個数	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の販売個数	前日, 前々日の2つ(当日は目的変数).	2 (1)
来客数	前日, 前々日の2つ(当日は未知).	2 (1)
ミスナールの体感温度	当日, 前日, 前々日の3つ.	3 (1)
降水量	当日, 前日, 前々日の3つ.	3 (1)
		計: 20(8)

表7 リンケの体感温度の説明変数(元気)

説明変数	備考	数
曜日		1 (1)
元気牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
元気牛乳の販売個数	前日, 前々日の2つ(当日は目的変数).	2 (1)
北海道牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の販売個数	当日, 前日, 前々日の3つ.	3 (1)
来客数	前日, 前々日の2つ(当日は未知).	2 (1)
リンケの体感温度	当日, 前日, 前々日の3つ.	3 (1)
降水量	当日, 前日, 前々日の3つ.	3 (1)
		計: 20(8)

表8 ミスナールの体感温度の説明変数(元気)

説明変数	備考	数
曜日		1 (1)
元気牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
元気牛乳の販売個数	前日, 前々日の2つ(当日は目的変数).	2 (1)
北海道牛乳の価格	当日, 前日, 前々日の3つ.	3 (1)
北海道牛乳の販売個数	当日, 前日, 前々日の3つ.	3 (1)
来客数	前日, 前々日の2つ(当日は未知).	2 (1)
ミスナールの体感温度	当日, 前日, 前々日の3つ.	3 (1)
降水量	当日, 前日, 前々日の3つ.	3 (1)
		計: 20(8)

6. 実験結果(1/4)

予測結果(北海道牛乳)

※表中の網掛けの数値は、その行で最も良い値。

表9 予測結果まとめ(北海道牛乳)

北海道牛乳 全期間 [1年]		気温		リンケの体感温度		ミスナールの体感温度	
		前日あり	前日なし	前日あり	前日なし	前日あり	前日なし
相関係数		0.926	0.947	0.938	0.961	0.931	0.956
誤差	平均	31	24	29	22	30	24
	最大値	167	283	161	115	192	172
	標準偏差	30	27	27	21	28	22
累計誤差		11019	8582	10279	8071	10765	8580

北海道牛乳は、リンケの体感温度の前日情報がないときに相関係数が最も高く、誤差の平均や最大値、標準偏差や累計誤差が最も小さい値になった。また、気温でも体感温度でも、前日情報がないときの方が、相関係数が高く、誤差が小さくなった。

6. 実験結果(2/4)

予測結果(北海道牛乳)

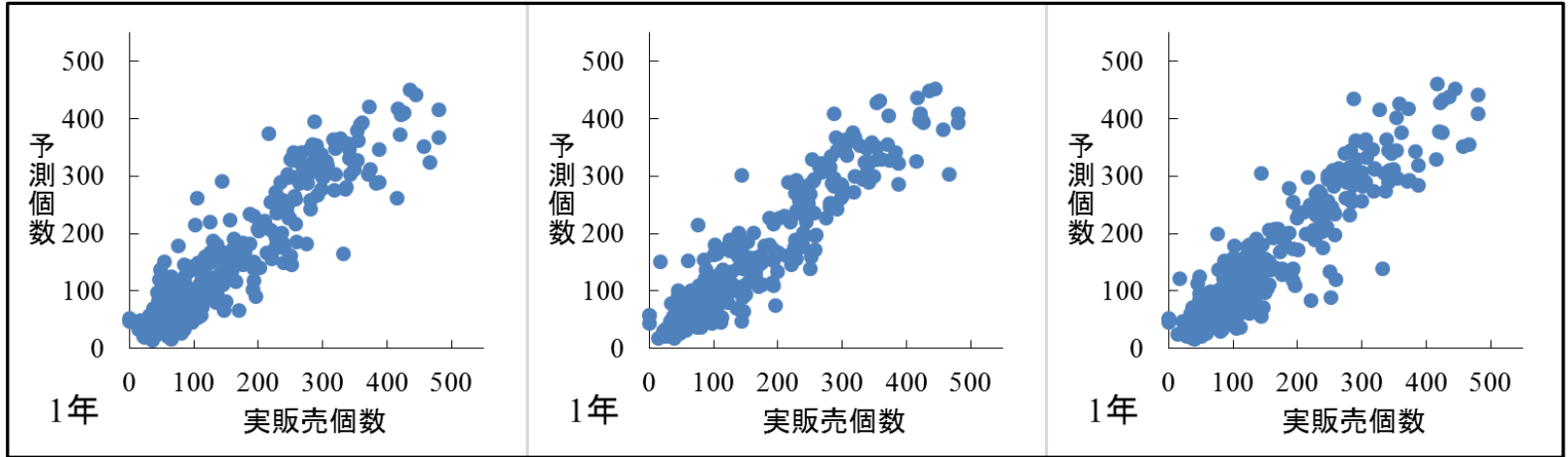


図2 実販売個数と予測個数の散布図(前日あり)

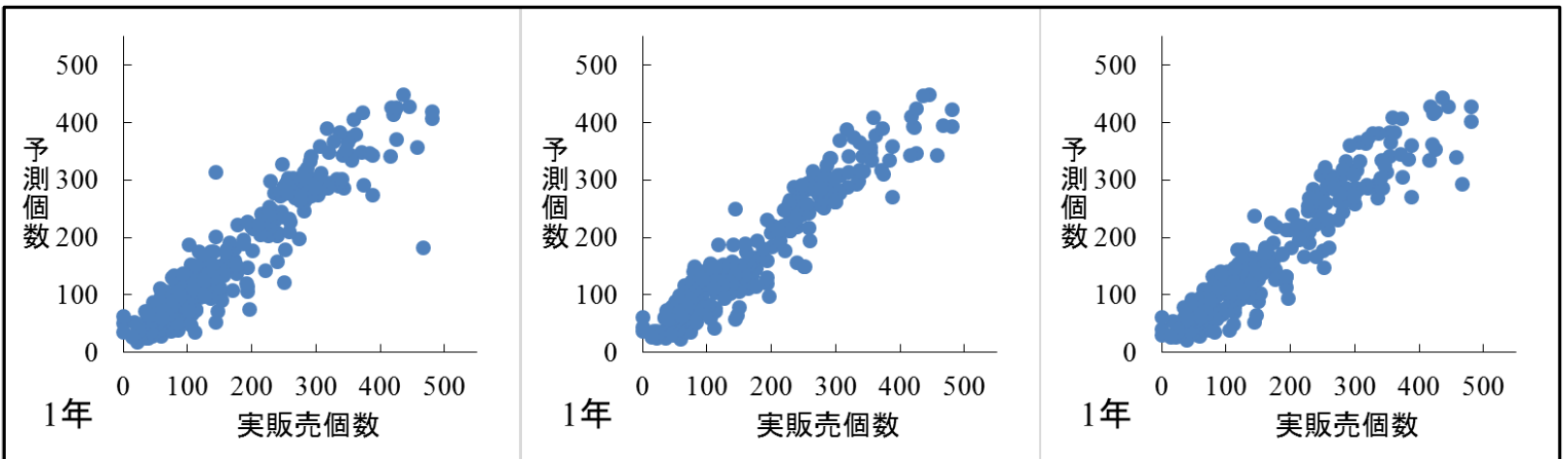


図3 実販売個数と予測個数の散布図(前日なし)

※それぞれ左から、気温、リンケの体感温度、
ミスナールの体感温度の散布図。

6. 実験結果(3/4)

予測結果(元気牛乳)

※表中の網掛けの数値は、その行で最も良い値.

表10 予測結果まとめ(元気牛乳)

元気牛乳 全期間 [1年]		気温		リンケの体感温度		ミスナールの体感温度	
		前日あり	前日なし	前日あり	前日なし	前日あり	前日なし
相関係数		0.957	0.971	0.961	0.972	0.957	0.974
誤差	平均	27	22	27	22	28	21
	最大値	212	162	169	157	169	156
	標準偏差	26	22	24	21	26	20
累計誤差		9863	7975	9687	7978	10102	7712

元気牛乳は、ミスナールの体感温度の前日情報がないときに相関係数が最も高く、誤差の平均や最大値、標準偏差や累計誤差が最も小さい値になっているが、他の変数の時の結果とほとんど差はない。こちら、北海道牛乳と同様に、前日情報がないときの方が予測結果が良い。

6. 実験結果(4/4)

予測結果(元気牛乳)

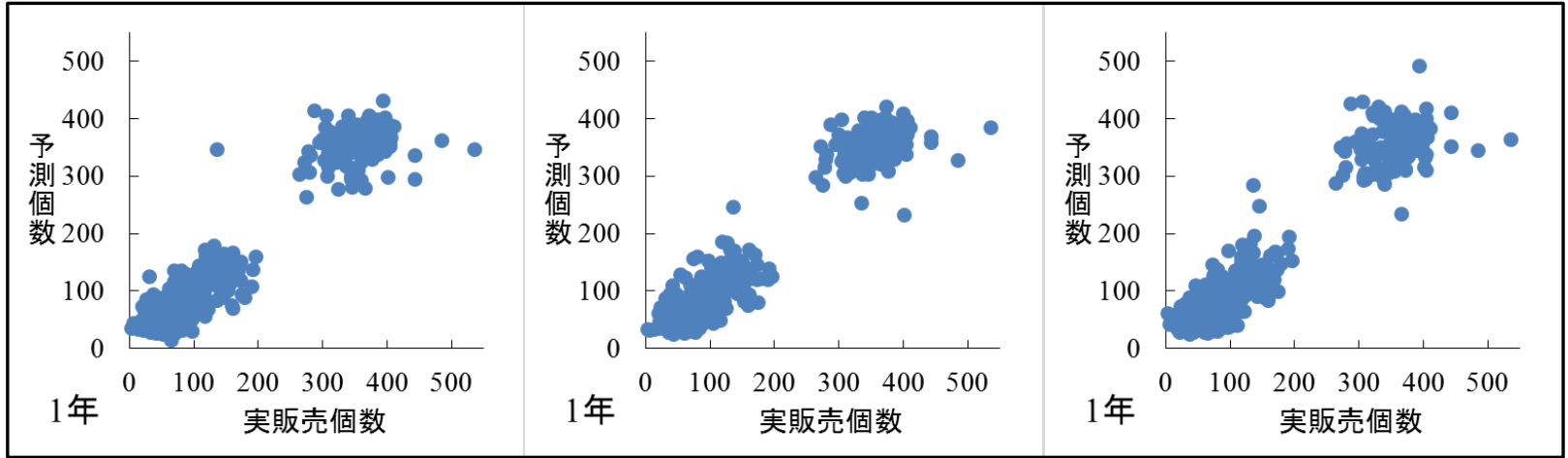


図4 実販売個数と予測個数の散布図(前日あり)

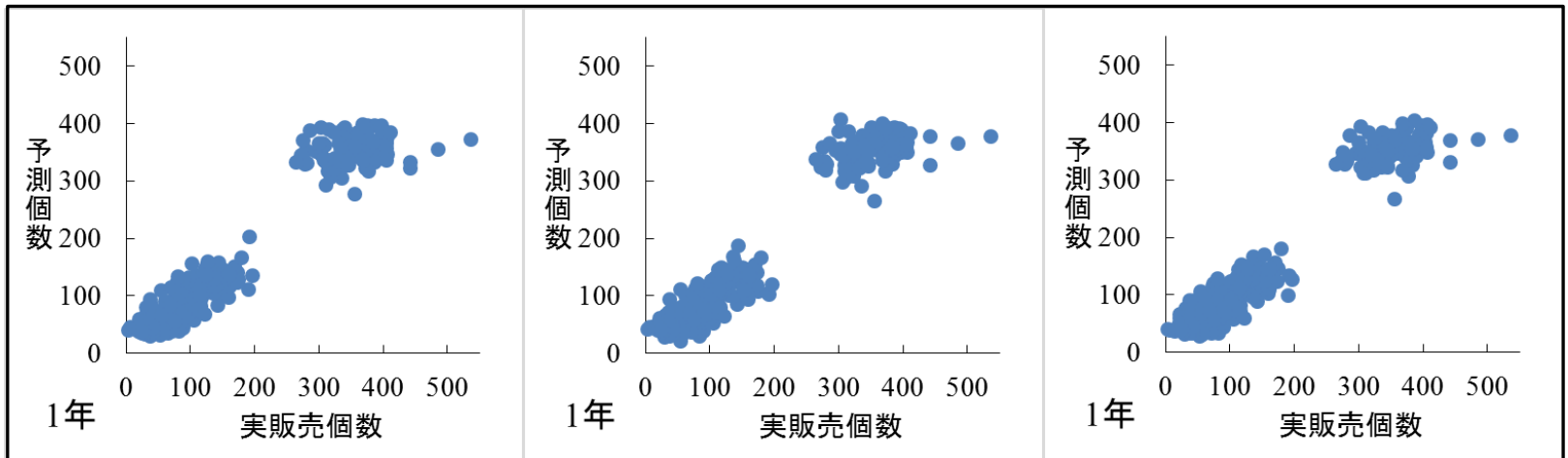


図5 実販売個数と予測個数の散布図(前日なし)

※それぞれ左から、気温、リンケの体感温度、ミスナールの体感温度の散布図。

7. 考察・まとめ(1/2)

ニューラルネットワークの説明変数に、
気温、リンケの体感温度、ミスナールの体感温度をそれぞれ用い、
北海道牛乳と元気牛乳の販売個数の予測を行った。

特に北海道牛乳で、気温よりも体感温度を用いた方が
相関係数が高く、誤差が小さくなるという結果を得られた。

またリンケの体感温度では、誤差の最大値を
ミスナールの体感温度の3分の2程度に抑えることができた。

実販売個数と予測個数の散布図を見ると、どちらの牛乳も
前日ありよりも 前日なしの方がばらつきが少なく、正の相関関係がみられる。

7. 考察・まとめ (2/2)

どちらの牛乳の予測結果でも、前日ありよりも **前日なし**の方が良いという結果になったが、これには、以下のような原因が考えられる。

- ・前日ありの説明変数の数が、前日なしよりも2倍以上多いため、過学習を起こしている。
- ・前日・前々日の変数がノイズとなっている。（「販売個数」の変数を例にすると、予測する日は定価でも、前日・前々日のいずれかが特価により販売個数が多い等。）

元気牛乳の散布図で、プロットが二つのグループに分かれているのは、今回の予測対象の店舗6では、元気牛乳の実販売個数のヒストグラムがふた山型(図6)になっているためである。

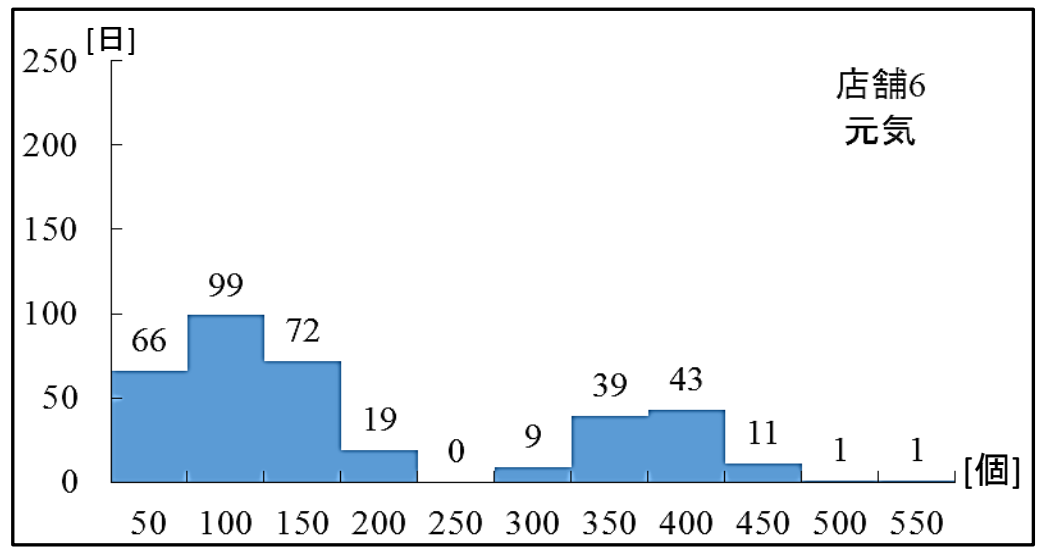


図6 実販売個数のヒストグラム

8. 今後の課題

- 暑いときよりも寒いときのほうが、人間にとって体感温度は敏感に反応する[13]のか、他店舗の予測をして月別または季節ごとに予測結果を比較する。
- 今回行った、説明変数に気温ではなく体感温度を用いる予測方法が、他店舗でも効果があるのか調査する。
- 変数が多いと過学習を起こしている可能性があるため、主成分分析をして、変数の数を減らした予測を行う。
- ニューラルネットワークの設定や、説明変数の選択について検討する。



9.参考文献(1/2)

- [1]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, 31(2), F-F13_1-8, 2016.
- [2]福島 和宏, 中谷 朋昭, 上田 雅夫, ”世界金融危機による消費者の食品購買行動の変化--日別POSデータによる分析”, 農経論叢, 66, pp13-19, 2011.
- [3]鈴木 督久, 小宮 葉子, ” POSデータのデータマイニング”, 日本SASユーザー会, vol.19, pp243-244, 2000.
- [4]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, 52(3) (通号 676) 2001.3, pp249-254, 2001.
- [5]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, 第16巻3号, pp15-23, 2009.
- [6]矢坂 雅充, ”乳業の構造”, 長期金融, 第72巻, pp78-115, 1991
- [7]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, vol.19, pp235-242, 2000 .

9.参考文献(2/2)

- [8]鈴木 督久, 小宮 葉子, ”POSデータのデータマイニング”, 日本SASユーザー会, vol.19, pp243-244, 2000 .
- [9] 2016年7月7日アクセス, ”ニューラルネットワーク”, <http://www.geocities.co.jp/SiliconValley-Cupertino/3384/nn/NN.html>.
- [10]高野 哲夫, ”ニューラルネットワークによるサクラ開花日の学習・予測実験 : 山形・新潟を例に(調査ノート)”, 天気 62(9), pp807-812, 2015.
- [11]山村 豊, 中村 和之, 山形 崇紘, ”ニューラルネットワークによる文字認識”, 函館工業高等専門学校紀要 47, pp53-58, 2013.
- [12]竹村 遼, 鈴木 達哉, 宮本 道子, 飯塚 佳代, ”外食産業の売上予測について: - 外食チェーンのある店舗における分析結果より - ”, 経営情報学会全国研究発表大会要旨集, 2012s(0), pp128-131, 2012.
- [13]緒方 知之(編著), ”鈴木敏文 考える経営”, 日本経済新聞社, 2005.

ご清聴ありがとうございました。