

変数選択によるメタ特徴を用いた 需要予測に関する研究

大同大学情報学部情報システム学科 齊藤哲雄
指導教員 柴田慎一

目次

- 研究背景
- 関連研究
- 研究目的
- 提案手法
- 使用するデータ
- 実験概要
- 実験結果
- 考察
- まとめ
- 今後の課題
- 参考文献

研究背景

機械学習において、データに対してどのアルゴリズムを使うべきかは、データの特徴や用途によって様々である。

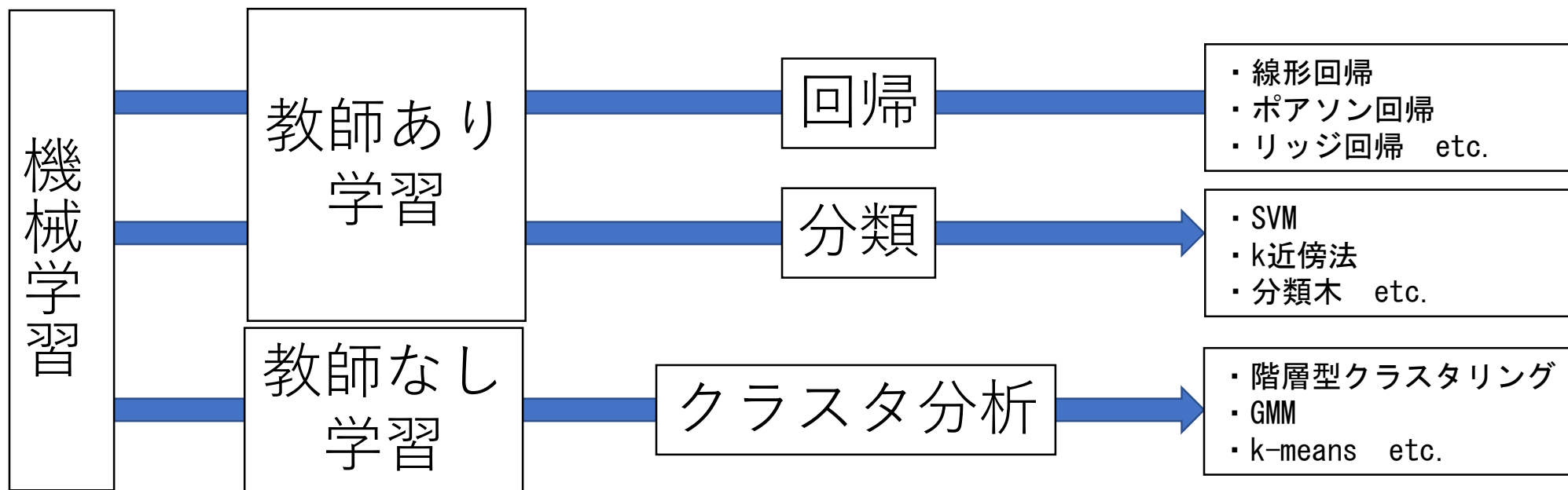


図1. 機械学習アルゴリズム

研究背景

機械学習の識別器は多数あり、
すべての問題に万能なアルゴリズムは存在しない[1]。
データセットに数多くある識別器を全て試すのは時間が掛かる。



最適な識別器を選択する手法として、**メタ学習**がある[2]。
メタ学習は、**メタ特徴**と呼ばれるデータセット自体が持つ特徴[1]を使って
データセットと最もよい識別器との関係を見つけることである。

メタ特徴の例として、インスタンス数、属性数、クラス数、
歪度、尖度、クラスエントロピーなどがある。

[1] 南保英孝[他], "メタ特徴の最適化処理による識別器構築アルゴリズム自動選択システム", 2016.

[2] B. Pavel, S. Carlos, "Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results", 2003.

関連研究

メタ特徴によるPOSデータの需要予測モデルの検討[3]

商品の販売量を予測するにあたり最適な学習モデルを検討した。POS (Point Of Sales) データのメタ特徴を抽出し最適な学習モデルを選択し、予測誤差の削減を目的とした。

結果： 当日の販売価格が定価、特価によって価格帯を分け予測した場合に予測精度が向上した結果が得られた。

課題： 説明変数の削減などを考慮し学習モデルを選択する必要がある。

研究目的

メタ特徴を用いて識別器選択システムの構築を行う。
学習データを変更し、最適な識別器を選択することができる
学習データを見つける。

メタ学習は機械学習の分類問題で扱われる。
それを需要予測といった回帰問題として扱い、
回帰問題でのメタ学習の有用性を示す。
識別器選択システムが予測誤差が小さくなる最適な識別器を
正しく選択させることを研究目的とする。

使用するデータ

使用するデータは、スーパーマーケット9店舗のPOSデータを扱う。
商品は購入頻度が高く、日配品である牛乳を使用する。
使用期間は1年間(2007年4月～2008年3月)である。

表1. スーパーマーケットの詳細表

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 (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)

使用するデータ

POSデータの中から、

「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」

の3種類の牛乳の販売量に関するデータを抽出し使用する。

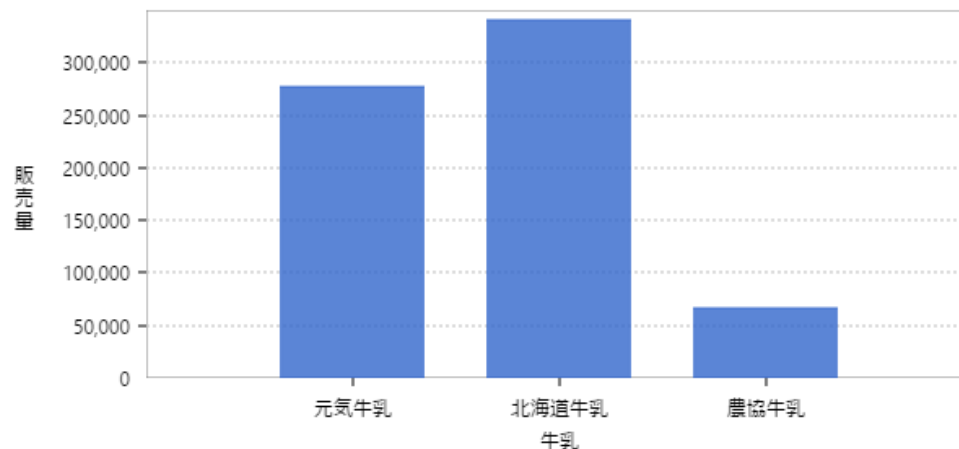


図2. 各牛乳の販売量

使用するデータ

全9店舗の元気牛乳と北海道牛乳の月平均販売量の
散布図を作成した。

散布図から本研究では、

- 販売量が平均的な店舗3
- 販売量が多い傾向の店舗6
- 販売量が少ない傾向の店舗8

上記の3店舗のデータを使用する。

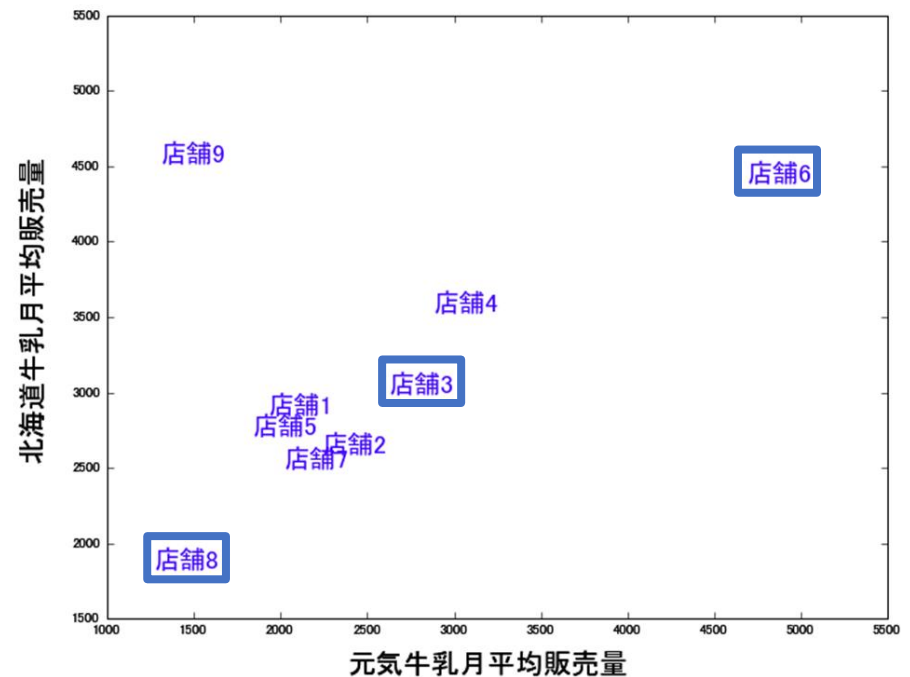


図3. 店舗別月平均販売量の散布図

提案手法

VMS (Visual Mining Studio)

VMSは株式会社NTTデータ数理システムが提供するデータマイニングツールである。データマイニングに必要なデータの前処理から分析、処理など高機能なツール群を誰でも簡単に利用することができる特徴がある。



VRP (Visual R Platform)

VRPはR言語が持つ多くの統計手法を、簡単なマウス操作で実行できる統計解析ツールである。



提案手法

使用するツール

VMS	VRP
Neural Network	Rグラフ表示
k-NN分析	
Radial Basis Function Network	
Support Vector Machine	
Naïve Bayes	
Decision Tree	
説明変数重要度	

提案手法

メタ特徴

データセットの内容ではなくデータセット自体が持つ特徴

Rのパッケージを使うことで
192種類のメタ特徴を抽出することができる。



R言語

Rは統計解析および統計グラフ作成のための開発実行環境

Rではmfeパッケージを用いてメタ特徴の抽出を行う。
mfeパッケージはバージョン3.5.0以上でないと使用できない。
そのため、メタ特徴抽出をVRPではなく、Rで行う。

提案手法

ニューラルネットワーク (Neural Network : NN)

人間の脳内にある神経回路網を人工ニューロンという数式的なモデルで表現したもの。学習能力に優れており、データ量が多いほど精度が良いモデルが構築可能である[4]。

$$\text{softmax}(x_k) = \frac{\exp(x_k)}{\sum_{k'} \exp(x_{k'})}$$

RBF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。放射基底関数はガウス関数がよく使われる。少数のデータに対して高い予測精度が構築可能である[5]。

$$\text{gaussian}(x) = \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

[4] 鈴木督久, “ニューラルネットワークによる牛乳販売量の予測”, 2001.

[5] 畠中利治, 近藤伸彦, 魚崎勝司, “多目的進化計算によるパレート最適なRBFネットワークの構成法”, 2005.

提案手法

k-近傍法 (k-Nearest Neighbor : k-nn)

データ間の距離が近いサンプルをk個を選び、回帰問題では平均化を、分類問題では多数決を取って分類する。

データ全体の分布に考慮せずに識別が可能である[6]。

$$Euclid(x, y) = \sum_i (x_i - y_i)^2 + \sum_{i,j} \left(\frac{C_{xij}}{C_x} - \frac{C_{yij}}{C_y} \right)^2$$

サポートベクターマシン (Support Vector Machine : SVM)

ニューロンの最も単純な線形しきい素子を拡張したパターン識別器である[7]。非線形の識別を扱えるものの局所解の問題がないという利点がある。

$$gaussian(x, y) = \exp\left(-\frac{(x - y)^2}{\sigma^2}\right)$$

[6] 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.

[7] 小野田崇, “知の科学 サポートベクターマシン”, 2007.

提案手法

ナীবベイズ (Naïve Bayes)

単純ベイズ法である。文書分類の確率論的手法の代表的な分類器である[8]。

$$p(c|x_1, \dots, x_n) = \frac{p(x_1, \dots, x_n|c)p(c)}{p(x_1, \dots, x_n)}$$

決定木 (Decision Tree)

特徴と得られる可能性のある結果の関係をモデリングするために木構造を使う分類器である[9]。

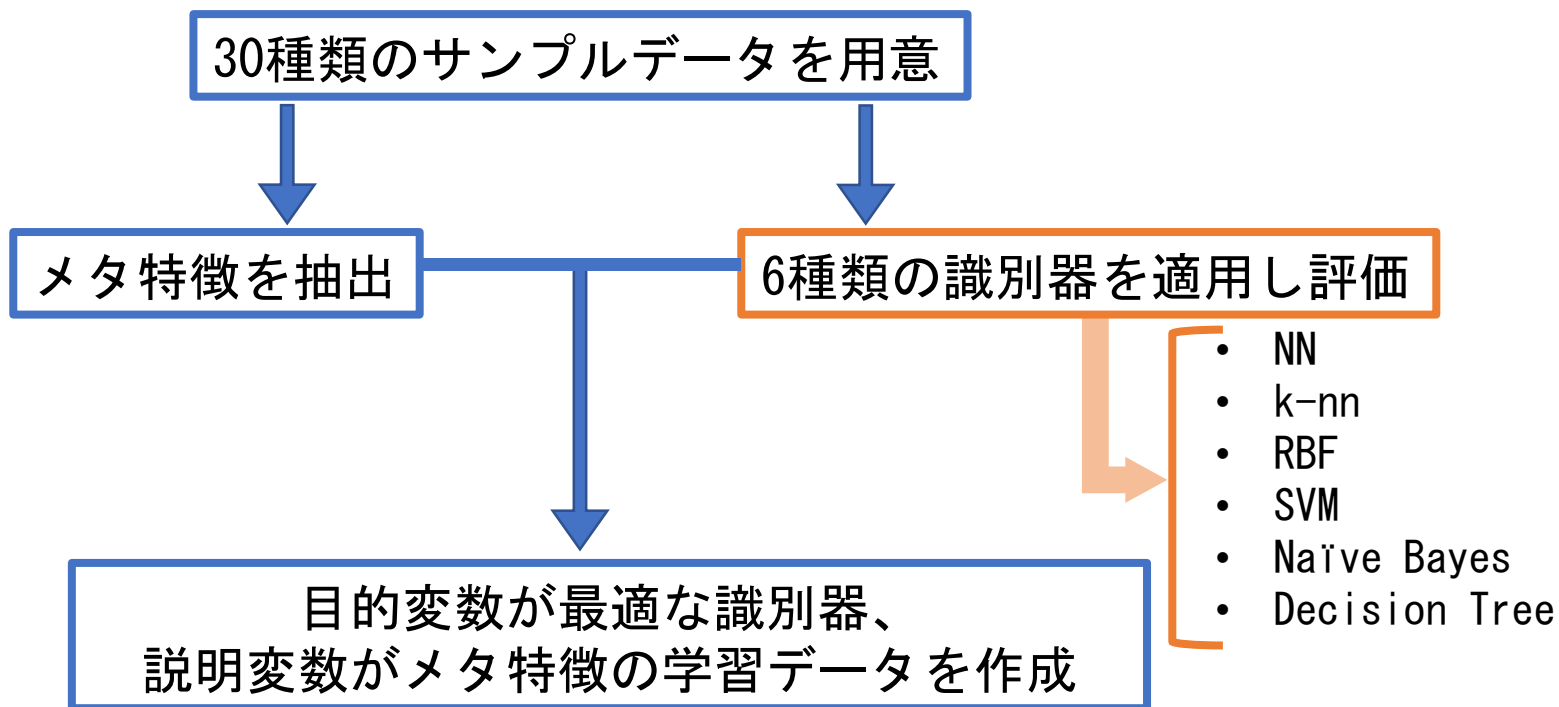
$$gainRatio(S, A) = \frac{gain(S, A)}{splitInfo(S, A)}$$

[8] 正田備也[他], “混合ディレクレ分布を用いた文書分類の精度について”, 2007.

[9] 長尾高弘, “Rによる機械学習”, 2017.

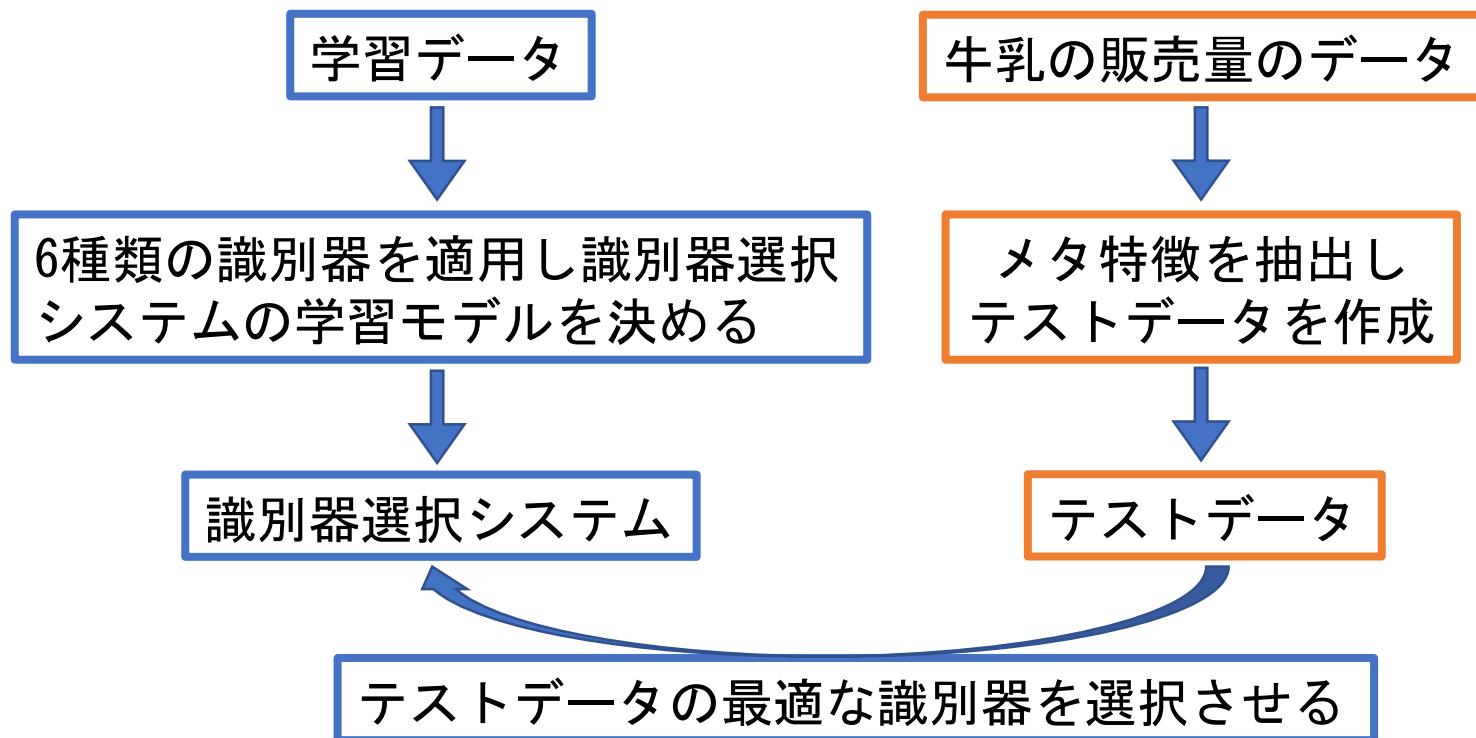
実験概要

識別器選択システムの構築のため、学習データの作成を行う。



実験概要

学習データから識別器選択システムを構築し、テストデータを入力して最適な識別器を選択させる。



実験概要

サンプルデータセットをWekaやUCI repositoryから30種類用意した。

表2. サンプルデータセット

No.	データ名	サンプル数	クラス数	No.	データ名	サンプル数	クラス数
1	がん患者	286	2	16	豆	119	8
2	糖尿病	768	2	17	コンクリート1	1030	9
3	アヤメ	150	3	18	インストラクター	647	13
4	アワビ	297	3	19	需要予測	60	20
5	コンタクトレンズ	24	3	20	コンクリート2	1030	21
6	卸売顧客	440	3	21	メタデータ	504	24
7	知識モデリング	258	4	22	葉データ	340	30
8	重量挙げ	366	5	23	CPU	209	32
9	ガラス	214	6	24	収益	997	32
10	車満足度	60	6	25	フェイスブック	495	34
11	セグメント	810	7	26	消費電力	1043	42
12	胸部手術	470	7	27	タンパク質	1143	42
13	国勢調査	693	7	28	ラスベガス	408	42
14	製品誤差	50	7	29	山火事	517	43
15	マウス	1080	8	30	超伝導	1063	52

実験概要

Rのmfeパッケージを使ってサンプルデータのメタ特徴を抽出する。

コード

```
> library(mfe)
> x <- read.csv("1_がん患者2.csv")
> metafeatures(Class~., x, groups = c("all"), summary=c("min", "max", "mean", "sd"))
```

出力結果

```
general.attrToInst      general.catToNum      general.freqClass.min  general.freqClass.max
      3.146853e-02              NA      2.972028e-01      7.027972e-01
general.freqClass.mean  general.freqClass.sd  general.instToAttr      general.nrAttr
      5.000000e-01      2.867986e-01      3.177778e+01      9.000000e+00
general.nrBin           general.nrCat          general.nrClass          general.nrInst
      2.000000e+00      9.000000e+00      2.000000e+00      2.860000e+02
      ⋮
      ⋮
```

実験概要

30種類のサンプルデータに対して識別器を適用して評価する。

- ニューラルネットワーク
- k近傍法
- RBFネットワーク
- サポートベクターマシン
- Naïve Bayes
- Decision Tree

サンプルデータ1つひとつに識別器を適用し交差検証を行う。
F値が大きくなる最適な識別器を求める。

実験概要

VMSで30種類のサンプルデータに6種類の識別器を交差検証を用いて適用した。

プロジェクト図を以下に示す。

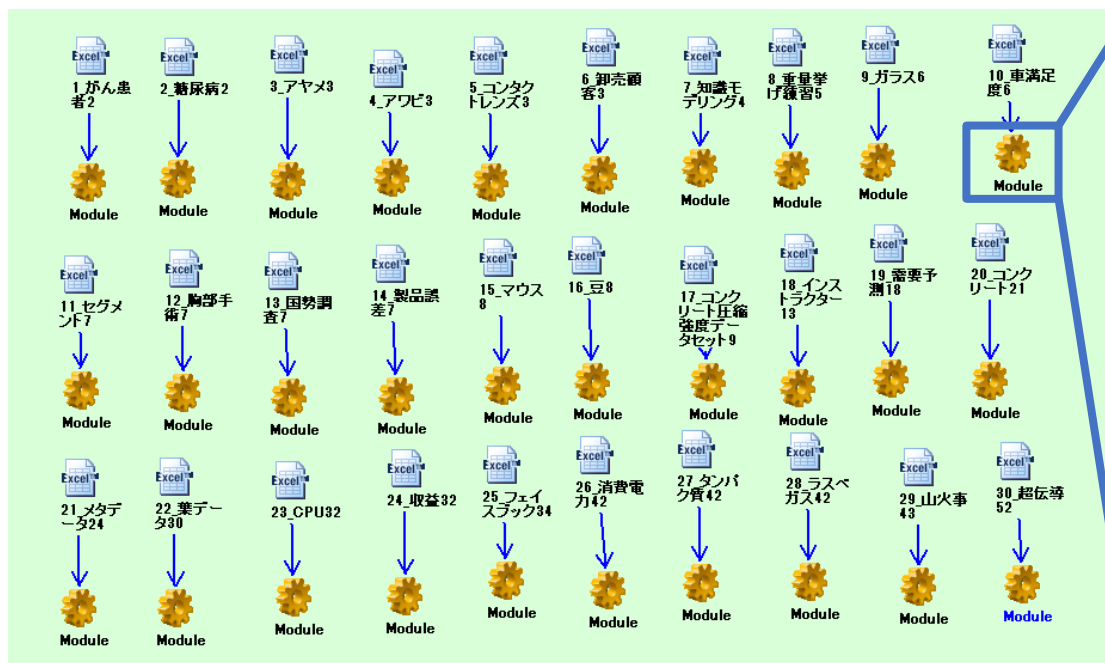


図4. プロジェクト図の一部

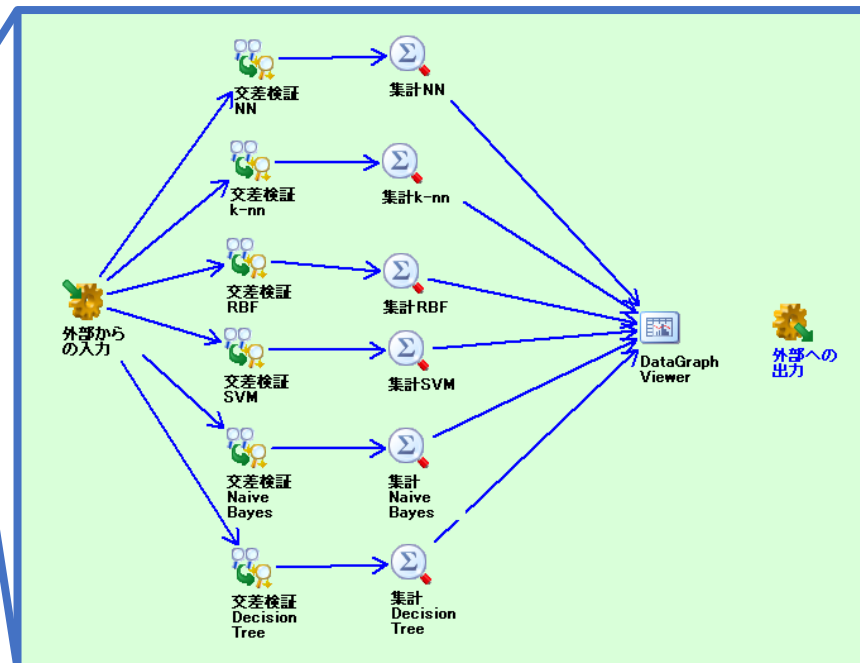


図5. モジュール内部図

実験概要

適用した識別器のパラメータ

表3. ニューラルネットのパラメータ

目的関数	Softmax-Multinomial
隠れ層のユニット数	データの説明変数の数
繰り返し最大数	1,000

表4. k-nnのパラメータ

距離関数	Euclid
k 値	5

表5. RBFネットワークのパラメータ

目的関数	Softmax-Multinomial
Weight Decay	0
係数	1.2
繰り返し最大数	1,000

表6. SVMのパラメータ

カーネル関数	Gaussian
パラメーター	0.1
Slack変数の係数	5

表7. Naïve Bayesのパラメータ

情報量基準	AIC
閾値	0
離散化方法	誤判別数

表8. Decision Treeのパラメータ

分岐方法	InfoGain Ratio
節点最小データ数(%)	0.8
節点の不純度	0.01
変数の最大分岐数	5

実験概要

メタ特徴と最適な識別器の結果から学習データを作成した。

表9. 学習データ

データ名	general.attrToInst	general.catToNum	general.freqClass.min	...	識別器
がん患者	0.031469	NA	0.297203	...	SVM
糖尿病	0.010417	0	0.348958	...	NN
アヤメ	0.026667	0	0.333333	...	SVM
⋮	⋮	⋮	⋮	⋮	⋮

説明変数をメタ特徴

目的変数を
最適な識別器

実験概要

学習データのうち、マウスデータと豆データは、最適な識別器が複数ある結果 (F値が1.0) となった。そこで、複数ある識別器を別々のデータとしたものと、マウスデータと豆データを削除したものをそれぞれ、**学習データ1** (データ数33)、**学習データ2** (データ数28) とした。

学習データ1

データ名	メタ特徴	識別器
⋮	⋮	⋮
製品誤差	⋯	SVM
マウス	⋯	k-nn
マウス	⋯	RBF
マウス	⋯	Decision Tree
豆	⋯	NN
豆	⋯	k-nn
コンクリート1	⋯	SVM
⋮	⋮	⋮



データ名	メタ特徴	識別器
⋮	⋮	⋮
製品誤差	⋯	SVM
マウス	⋯	k-nn,RBF,Decision Tree
豆	⋯	NN,k-nn
コンクリート1	⋯	SVM
⋮	⋮	⋮



学習データ2

データ名	メタ特徴	識別器
⋮	⋮	⋮
製品誤差	⋯	SVM
コンクリート1	⋯	SVM
⋮	⋮	⋮

実験概要

次に、識別器選択システムの学習モデルを決める。
作成した学習データに6種類の識別器で交差検証を行い、
F値が大きいモデルを決定する。

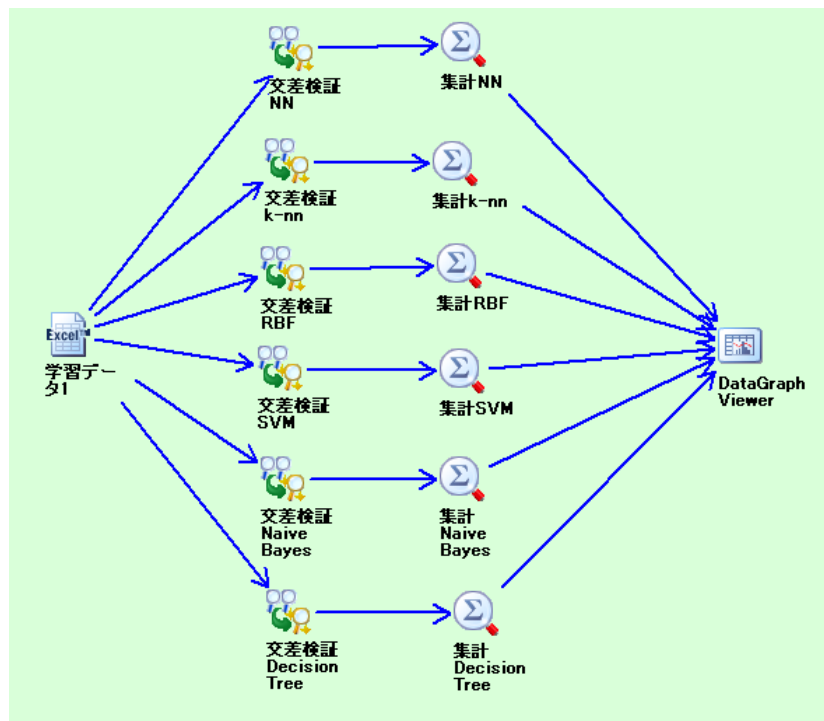


図6. 学習データの最適な学習モデルを決めるプロジェクト図

実験概要

6種類の識別器で交差検証を行い、下記の結果が得られた。

	NN	k-nn	RBF	SVM	Naïve Bayes	Decision Tree
F値	0.24	0.18	0.25	0.06	0.07	0.26

以上の結果から、識別器選択システムの学習モデルは Decision Tree が良いと判断した。

実験概要

テストデータを作成する。
牛乳の販売量のデータを示す。

表10. 牛乳の販売量のデータ

説明変数	備考	数
曜日	日曜～土曜までを2値で表現	7
来客数	前日,前々日	2
平均気温	当日,前日,前々日	3
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
牛乳の販売個数	前日,前々日(当日は目的変数)	2
牛乳の販売単価	当日,前日,前々日	3
説明変数の合計		26

上記のデータをRでメタ特徴の抽出を行い、テストデータを作成した。また、上記のデータを6種類の識別器で需要予測を行い、観測値と予測値の累積誤差を求めた。累積誤差が少ない結果となる識別器をシステムが正しく選択できるか、検証する

実験概要

店舗別に6種類の識別器を使って累積誤差を求めた。

表11. 店舗3の牛乳の販売量の累積誤差

	元気牛乳						北海道牛乳						農協牛乳					
	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree
累積誤差	2,375	4,761	2,103	1,117	3,475	3,617	2,844	5,214	1,388	1,058	4,008	4,280	994	1,350	1,178	950	1,246	1,160
平均値	7	13	6	3	10	10	8	15	4	3	11	12	3	4	3	3	4	3
標準偏差	10	19	9	3	14	18	15	20	5	6	20	21	2	5	2	3	3	4
最大値	67	205	67	30	101	182	118	118	39	94	139	129	11	70	11	42	24	55

表12. 店舗6の牛乳の販売量の累積誤差

	元気牛乳						北海道牛乳						農協牛乳					
	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree
累積誤差	3,425	8,533	2,177	1,427	4,827	6,315	3,292	8,744	1,450	1,538	4,765	6,251	1,213	1,385	1,227	905	1,291	1,115
平均値	9	24	6	4	13	17	9	25	4	4	13	18	3	4	3	3	4	3
標準偏差	15	30	11	20	21	27	19	34	7	17	24	28	3	6	3	2	3	3
最大値	102	260	102	366	129	210	126	215	54	272	151	160	14	95	15	9	15	39

表13. 店舗8の牛乳の販売量の累積誤差

	元気牛乳						北海道牛乳						農協牛乳					
	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree	NN	k-nn	RBF	SVM	NaiveBayes	DecisionTree
累積誤差	1,520	3,240	1,614	864	3,490	2,518	1,356	3,436	1,708	958	2,768	948	1,054	1,278	1,130	886	1,324	1,240
平均値	5	10	5	3	10	7	4	10	5	3	8	3	3	4	3	2	4	3
標準偏差	7	13	6	2	14	13	4	12	6	3	11	2	2	4	2	1	4	4
最大値	71	91	49	28	78	118	32	80	41	43	76	18	16	51	16	5	33	44

実験概要

識別器選択システムでテストデータの最適な識別器を選択させる。
下記にプロジェクト図を示す。

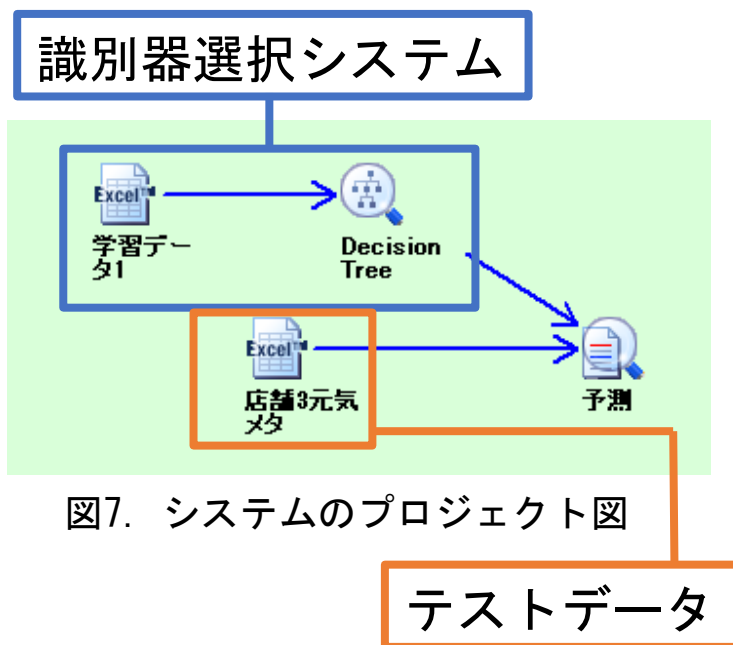


図7. システムのプロジェクト図

実験概要

精度や汎化性能の向上のため、学習データを変数選択する。
VMS上で行える変数選択手法として次の2つがある。

Feature Selection	情報量基準を比較し目的変数と説明変数の関連性を探す。
説明変数重要度	Extra Treeを用いて説明変数の重要度を算出する。

今回は、学習が高速でランダム性により、教師データが少なくても過学習しにくい説明変数重要度を使う。

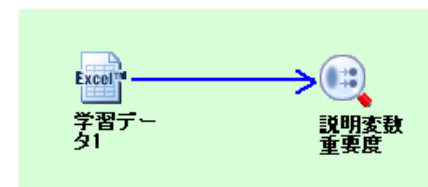


図8. 変数選択のプロジェクト図

重要度の平均は1となるように規格化されているため、重要度が1以上の説明変数を取り扱う。

実験結果

6つの識別器

- 実験1…学習データ1
- 実験2…学習データ2
- 実験3…学習データ1 + 変数選択
- 実験4…学習データ2 + 変数選択

5つの識別器 (Naïve Bayesを除いた)

- 実験5…学習データ3
- 実験6…学習データ4
- 実験7…学習データ3 + 変数選択
- 実験8…学習データ4 + 変数選択

実験結果 実験1 学習データ1

識別器を6つとした時の店舗3・6・8のシステム選択結果を示す。
識別器の()の数字は累積誤差少ない順番を表している。

表14. 店舗3

	識別器
元気牛乳	NaiveBayes(4)
北海道牛乳	RBF(2)
農協牛乳	DecisionTree(3)

表15. 店舗6

	識別器
元気牛乳	NaiveBayes(4)
北海道牛乳	k-nn(6)
農協牛乳	NaiveBayes(5)

表16. 店舗8

	識別器
元気牛乳	k-nn(5)
北海道牛乳	RBF(4)
農協牛乳	DecisionTree(4)

実験1の結果から、**Naïve Bayes**と出力される箇所が多いことがわかる。そこで、学習データの識別器の数をNaïve Bayesを除いた5種類で作成し、比較を行う。

また、複数選択された識別器があるため、**学習データ3**(33データ)、**学習データ4**(28データ)とする。

実験結果 実験2 学習データ2

学習データ2のシステム選択結果を以下に示す。

この先の実験2～8の累積誤差の比較は実験1と比較した値である。

表17. 店舗3

	識別器	累積誤差の比較
元気牛乳	SVM	-2358
北海道牛乳	NN	1456
農協牛乳	NaiveBayes	-68

表18. 店舗6

	識別器	累積誤差の比較
元気牛乳	NaiveBayes	0
北海道牛乳	NaiveBayes	-3979
農協牛乳	NaiveBayes	0

表19. 店舗8

	識別器	累積誤差の比較
元気牛乳	NaiveBayes	250
北海道牛乳	NN	-352
農協牛乳	NaiveBayes	84

累積誤差の合計が実験1より4,967減った。

実験結果 実験3 学習データ1 + 変数選択

変数選択をした学習データ1のシステム選択結果を以下に示す。

表20. 店舗3

	識別器	累積誤差の比較
元気牛乳	RBF	-1372
北海道牛乳	NN	1456
農協牛乳	SVM	-296

表21. 店舗6

	識別器	累積誤差の比較
元気牛乳	NN	-1402
北海道牛乳	NN	-5452
農協牛乳	SVM	-386

表22. 店舗8

	識別器	累積誤差の比較
元気牛乳	RBF	-1626
北海道牛乳	RBF	0
農協牛乳	SVM	-354

累積誤差の合計が実験1より9,432減った。

実験結果 実験4 学習データ2 + 変数選択

変数選択をした学習データ2のシステム選択結果を以下に示す。

表23. 店舗3

	識別器	累積誤差の比較
元気牛乳	RBF	-1372
北海道牛乳	RBF	0
農協牛乳	NN	-252

表24. 店舗6

	識別器	累積誤差の比較
元気牛乳	NN	-1402
北海道牛乳	NN	-5452
農協牛乳	NN	-78

表25. 店舗8

	識別器	累積誤差の比較
元気牛乳	RBF	-1626
北海道牛乳	NN	-352
農協牛乳	NN	-186

累積誤差の合計が実験1より10,720減った。

実験結果 実験5 学習データ3

学習データ3のシステム選択結果を以下に示す。

表26. 店舗3

	識別器	累積誤差の比較
元気牛乳	NN	-1100
北海道牛乳	SVM	-330
農協牛乳	NN	-252

表27. 店舗6

	識別器	累積誤差の比較
元気牛乳	SVM	-3400
北海道牛乳	SVM	-7206
農協牛乳	NN	-78

表28. 店舗8

	識別器	累積誤差の比較
元気牛乳	NN	-1720
北海道牛乳	SVM	-750
農協牛乳	NN	-186

累積誤差の合計が実験1より15,022減った。

実験結果 実験6 学習データ4

学習データ4のシステム選択結果を以下に示す。

表29. 店舗3

	識別器	累積誤差の比較
元気牛乳	NN	-1100
北海道牛乳	SVM	-330
農協牛乳	NN	-252

表30. 店舗6

	識別器	累積誤差の比較
元気牛乳	SVM	-3400
北海道牛乳	SVM	-7206
農協牛乳	NN	-78

表31. 店舗8

	識別器	累積誤差の比較
元気牛乳	NN	-1720
北海道牛乳	SVM	-750
農協牛乳	NN	-186

累積誤差の合計が実験1より15,022減った。

実験結果 実験7 学習データ3 + 変数選択

変数選択をした学習データ3のシステム選択結果を以下に示す。

表32. 店舗3

	識別器	累積誤差の比較
元気牛乳	NN	-1100
北海道牛乳	SVM	-330
農協牛乳	NN	-252

表33. 店舗6

	識別器	累積誤差の比較
元気牛乳	SVM	-3400
北海道牛乳	SVM	-7206
農協牛乳	NN	-78

表34. 店舗8

	識別器	累積誤差の比較
元気牛乳	NN	-1720
北海道牛乳	SVM	-750
農協牛乳	NN	-186

累積誤差の合計が実験1より15,022減った。

実験5, 6, 7は選択された識別器が同じであるため、

累積誤差の比較の値も同じ結果となった。

実験結果 実験8 学習データ4 + 変数選択

変数選択をした学習データ4のシステム選択結果を以下に示す。

表35. 店舗3

	識別器	累積誤差の比較
元気牛乳	RBF	-1372
北海道牛乳	RBF	0
農協牛乳	SVM	-296

表36. 店舗6

	識別器	累積誤差の比較
元気牛乳	RBF	-2650
北海道牛乳	RBF	-7294
農協牛乳	NN	-78

表37. 店舗8

	識別器	累積誤差の比較
元気牛乳	RBF	-1626
北海道牛乳	RBF	0
農協牛乳	SVM	-354

累積誤差の合計が実験1より13,670減った。

考察

- 実験5, 6, 7は同じ結果となった。
また、実験5, 6, 7が全店舗ともに、一番累積誤差を減らす結果となった。
- 識別器選択システムの学習データはNaïve Bayesを除いた識別器5つの学習データがより牛乳の販売量の予測誤差が少ない識別器を出力することが分かった。Naïve Bayesを目的変数とするメタ特徴が選択システムの誤りを引き起こす原因であると考えられる。
- 実験3, 4, 7, 8から、変数選択してもシステムの出力結果が変わらない、もしくは悪くなった。このことから、識別器選択システムは過学習を起こしていないと考えられる。

まとめ

たくさんある機械学習の識別器をデータセットのメタ特徴から、どの識別器が最適なモデルかをみつける手法としてメタ学習がある。30種類のサンプルデータを使って学習データを作成し、識別器選択システムの構築を行った。そして、牛乳の販売量のテストデータの最適な識別器を選択した。

学習データの最適な識別器の数を減らすと、予測誤差が減る結果となった。

また、変数選択を行った学習データで学習しても予測誤差は減らなかった。

今後の課題

- 識別器のパラメータはデフォルトであったので、パラメータを変更して選択システムの精度を上げる
- R以外でメタ特徴抽出を行えるソフトと比較を行う
- 牛乳の販売量のデータを定価と特価の価格帯に分けて予測を行う

参考文献

- [1] 南保英孝, 大塚敦史, 木村春彦, “メタ特徴の最適化処理による識別器構築アルゴリズム自動選択システム”, 科学・技術研究会, Vol. 5, No. 2, pp. 179–184, 2016.
- [2] B. Pavel, S. Carlos, “Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results”, Machine Learning, 50, pp. 251–277, 2003.
- [3] 守田真矢, 柴田慎一, 田嶋拓也, 阿部武彦, 木村春彦, “メタ特徴によるPOSデータの需要予測モデルの検討”, 2019.
- [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, Vol. 52, No. 3, pp. 249–254, 2001.
- [5] 畠中利治, 近藤伸彦, 魚崎勝司, “多目的進化計算によるパレート最適なRBFネットワークの構成法”, 2005.
- [6] 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”, 研究報告自然言語処理(NL), Vol. 2013-NL-211, No. 13, pp1–7, 2013.
- [7] 小野田崇, “知の科学 サポートベクターマシン”, 2007.
- [8] 正田備也, 高杉淳宏, 安達淳, “混合ディレクレ分布を用いた文書分類の精度について”, 情報処理学会論文誌, Vol. 48, No. SIG11, 2007.
- [9] 長尾高弘, “Rによる機械学習”, 2017.

ご清聴ありがとうございました

謝辞

VMS、VRPの簡単な操作性と多様な分析ツールにより、スムーズに研究を進めることができました。
提供していただきましたNTTデータ数理システムの皆様と、日頃よりご指導いただきました柴田慎一先生に心より感謝申し上げます。