

進化型多目的ファジィルール選択を用いた設計データからの知識獲得

大阪府立大学 大学院工学研究科
電気・情報系専攻 知能情報工学分野
桑島 功

はじめに

目的

多目的設計データからデータマイニング手法を用いて何らかの知識を獲得する.

多目的設計データ

パターン数:149

属性数:13

欠落値なし

設計個体	翼の形状パラメータ断面1 Eta=0.123(SOB)					翼の形状パラメータ断面2 Eta=0.294					空力抵抗	衝撃波強さ	翼構造重量
unit no.	XmaxL	maxL	XmaxTC	maxTC	前桁高さ	XmaxL	maxL	XmaxTC	maxTC	前桁高さ	CD	-Cp min	Weight [kg]
1	41.509	-9.002	36.299	15.098	11.871	49.069	-7.381	43.887	13.102	10.379	0.019	0.829	868.290
2	33.034	-9.675	31.277	16.041	12.732	42.919	-7.985	40.752	13.799	11.412	0.019	0.873	867.270
3	28.573	-9.629	28.067	16.085	12.889	34.968	-6.387	35.162	12.244	10.126	0.020	1.138	874.880
4	42.217	-9.890	38.373	15.908	11.884	31.494	-7.834	33.292	13.678	12.542	0.020	1.092	864.140
5	31.180	-8.892	28.828	15.318	13.575	39.359	-7.224	38.853	13.068	10.520	0.018	0.638	854.950
6	26.169	-9.842	25.803	16.320	14.731	20.739	-7.209	25.822	12.872	12.433	0.022	1.524	868.130
7	35.679	-9.176	29.816	15.542	13.754	49.106	-7.341	44.100	13.060	10.917	0.019	0.737	858.400
8	32.991	-8.695	30.447	15.076	12.993	32.525	-8.023	33.186	13.870	12.006	0.020	1.131	873.100
:	:	:	:	:	:	:	:	:	:	:	:	:	:

3

データの基本的分析

- 記述統計(最大, 最小, 標準偏差等)
- 属性間の相関関係の分析
- プロット(ヒストグラム, 散布図)
- データの平滑化

4

記述統計によるデータの特徴

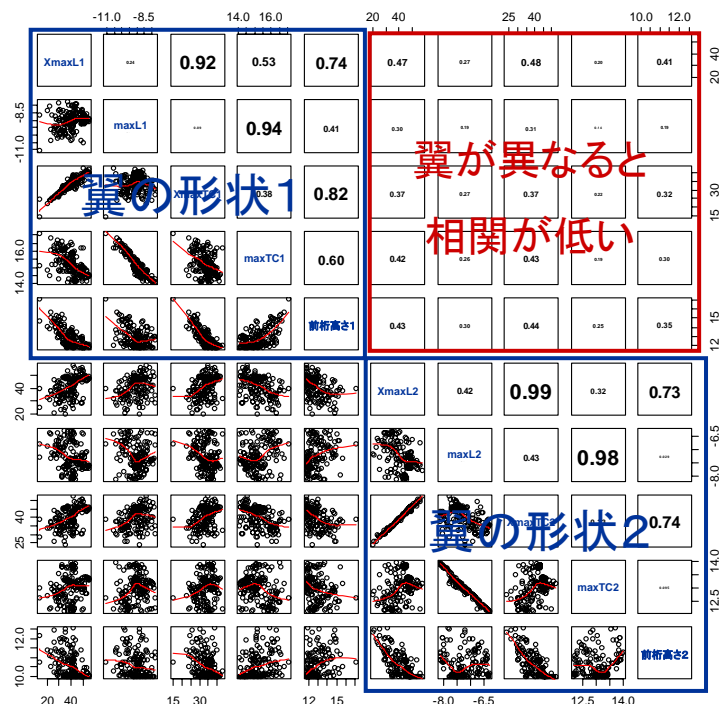
	属性名	値の種類	最大	最小	平均	標準偏差
翼の形状 パラメータ 断面1	XmaxL	実数値	54.981	13.619	39.356	8.297
	maxL	実数値	-7.767	-11.062	-9.078	0.631
	XmaxTC	実数値	43.062	14.630	33.304	4.668
	maxTC	実数値	17.095	14.083	15.238	0.695
	前桁高さ	実数値	17.047	11.740	12.845	1.042
翼の形状 パラメータ 断面2	XmaxL	実数値	58.119	20.291	41.123	8.255
	maxL	実数値	-6.295	-8.130	-7.244	0.504
	XmaxTC	実数値	52.488	23.516	39.893	6.102
	maxTC	実数値	13.971	12.153	13.012	0.484
	前桁高さ	実数値	12.542	9.964	10.607	0.671
目的関数値	空力抵抗 (CD)	実数値	0.024	0.018	0.019	0.001
	衝撃波強さ (-Cp min)	実数値	1.563	0.344	0.792	0.352
	翼構造重量 (Weight)	実数値	897.260	756.990	824.073	48.647

5

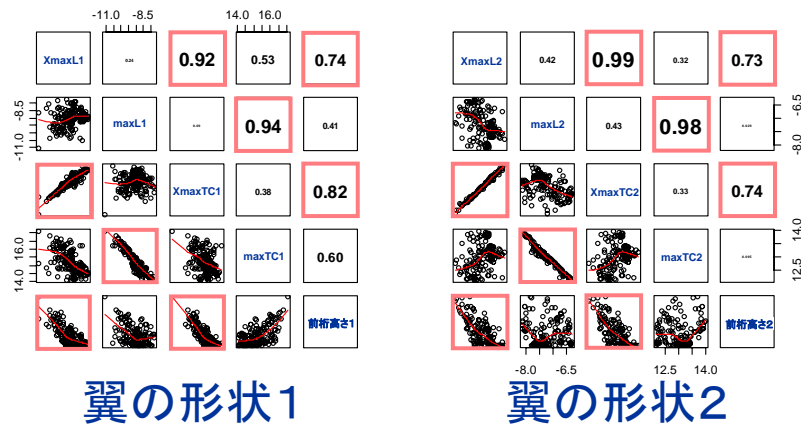
Pair plot

右上側:
相関係数

左下側:
散布図,
平滑化曲線



翼の形状ごとに分析



翼の形状1と2で傾向は似ている。

7

○ 相関ルールマイニング

相関ルールとは,

相関ルールの形式

$$A \Rightarrow B$$

A が存在するならば同時に B も存在する

データの属性間での依存関係を
記述する手法の一つ

9

使用例

マーケット・バスケット分析

$$\{\text{beer}\} \rightarrow \{\text{diaper}\}$$

$$\{\text{paper, envelopes}\} \rightarrow \{\text{stamps}\}$$

どの商品を同じバスケットの中に
入れるかを分析する

10

使用例: 2

数値属性相関ルール

Quantitative association rules

年齢が[30, 40] → 収入が[\$50000, \$70000]

区間によって属性値を定める

11

ファジィ集合を用いた相関関係

ファジィ相関ルール

Fuzzy association rules

年齢が“高い” → 収入が“多い”

区間の代わりにファジィ集合を用いる

12

ファジィ集合を用いることの利点

- 区間の閾値の望ましくない効果を避ける.
区間の場合には, 閾値を少しでも外れると結果が変わる.
- 数値的表現を言語的表現に変える.
“高い”, “低い”, “大きい”, “小さい”等.

13

Confidence & Support

- Confidence (信頼度)

$$c(A_q \Rightarrow B_q) = \frac{\sum_{p=1}^m \mu_{A_q}(x_p) \times \mu_{B_q}(x_p)}{\sum_{p=1}^m \mu_{A_q}(x_p)}$$

- Support (支持度)

$$s(A_q \Rightarrow B_q) = \frac{\sum_{p=1}^m \mu_{A_q}(x_p) \times \mu_{B_q}(x_p)}{m}$$

14

Confidence

区間の場合

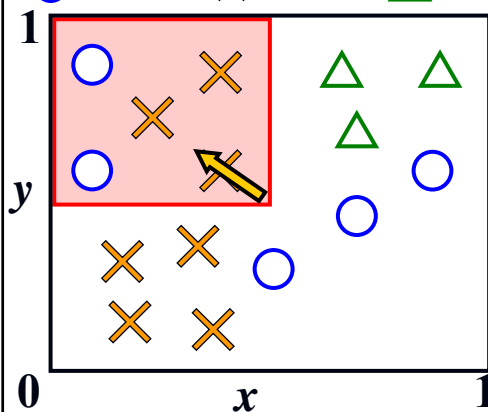
$$\text{Confidence} = \frac{\text{ルールが正答するパターン数}}{\text{条件部に一致するパターン数}}$$

15

Confidence

Rule: $x < 0.5$ and $y > 0.5 \Rightarrow$ **Class 1**

○ Class 1 × Class 2 △ Class 3



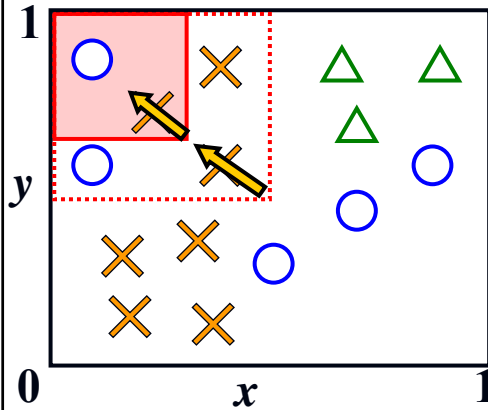
$$\begin{aligned} \text{Confidence} &= \frac{2}{5} \\ &= 0.4 \end{aligned}$$

16

Confidence

Rule: $x < 0.3$ and $y > 0.7 \Rightarrow$ Class 1

○ Class 1 × Class 2 △ Class 3



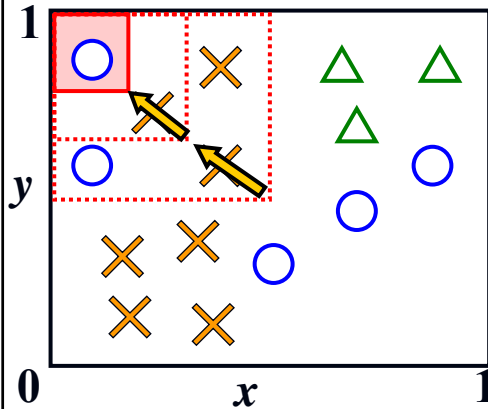
$$\text{Confidence} = \frac{1}{2} = 0.5$$

17

Confidence

Rule: $x < 0.1$ and $y > 0.9 \Rightarrow$ Class 1

○ Class 1 × Class 2 △ Class 3



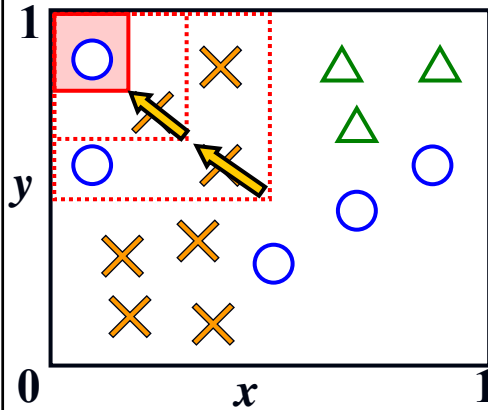
$$\text{Confidence} = \frac{1}{1} = 1.0$$

18

Confidence

Rule: $x < 0.1$ and $y > 0.9 \Rightarrow$ Class 1

○ Class 1 × Class 2 △ Class 3



Confidence

前件部の範囲が狭い
ルールほど値が高い
傾向がある。

19

Support

区間の場合

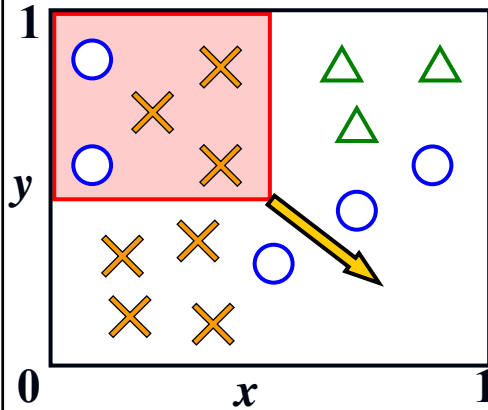
$$\text{Support} = \frac{\text{ルールが正答するパターン数}}{\text{全パターン数}}$$

20

Support

Rule: $x < 0.5$ and $y > 0.5 \Rightarrow$ Class 1

○ Class 1 × Class 2 △ Class 3



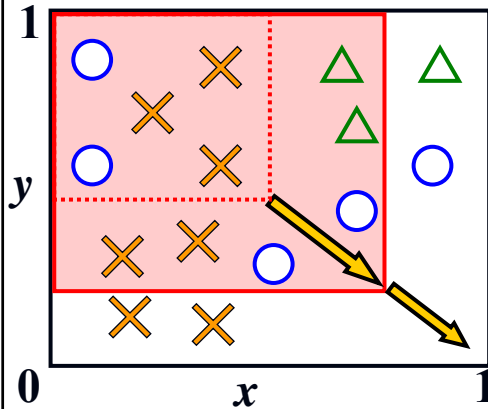
$$\text{Support} = \frac{2}{15} = 0.13$$

21

Support

Rule: $x < 0.7$ and $y > 0.3 \Rightarrow$ Class 1

○ Class 1 × Class 2 △ Class 3



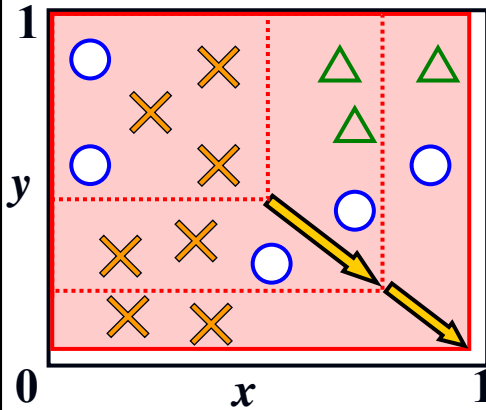
$$\text{Support} = \frac{4}{15} = 0.26$$

22

Support

Rule: $x < 0.9$ and $y > 0.1 \Rightarrow$ Class 1

○ Class 1 × Class 2 △ Class 3



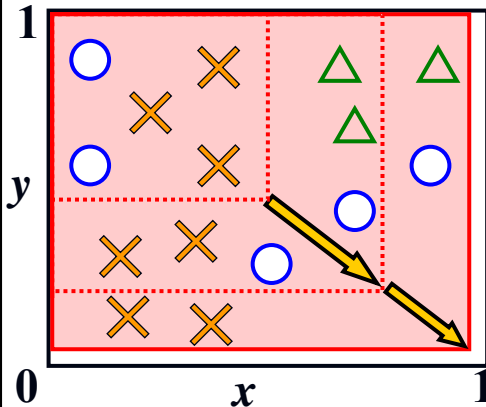
$$\text{Support} = \frac{5}{15} = 0.3$$

23

Support

Rule: $x < 0.9$ and $y > 0.1 \Rightarrow$ Class 1

○ Class 1 × Class 2 △ Class 3



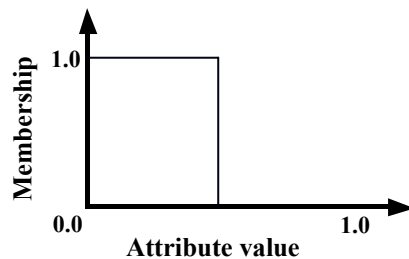
Support

前件部の範囲が広い
ルールほど値が高い
傾向がある。

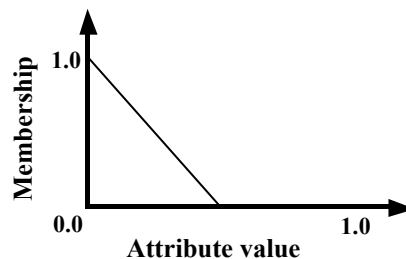
24

Confidence & Support

ファジィ集合の場合には、パターン数の代わりにメンバシップ関数値を用いる。



区間の場合



ファジィ集合の場合

25

相関ルールマイニング手法

マイニング手法の流れ

APRIORI (Agrawal et al., 1994)



FP-growth (Han et al., 2004)



TFP (Coenen et al., 2004)



⋮

26

相関ルールマイニング手法

目的

データに潜在する事象間の関係をルール形式の知識として獲得する.

$$A \Rightarrow B$$

A が存在するならば同時に B も存在する

27

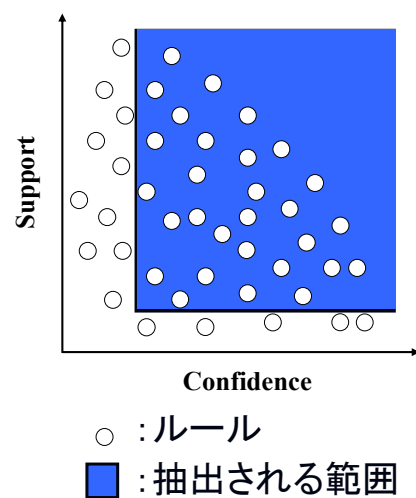
相関ルールマイニング手法

特徴

評価基準の閾値を満たすルールを全て抽出する.

評価基準:

- Confidence
- Support



28

数値実験

ファジィ相関ルールマイニングにより、
ファジィルールの形で知識の獲得を行う。

- Confidenceの高いルール
- Supportの高いルール

29

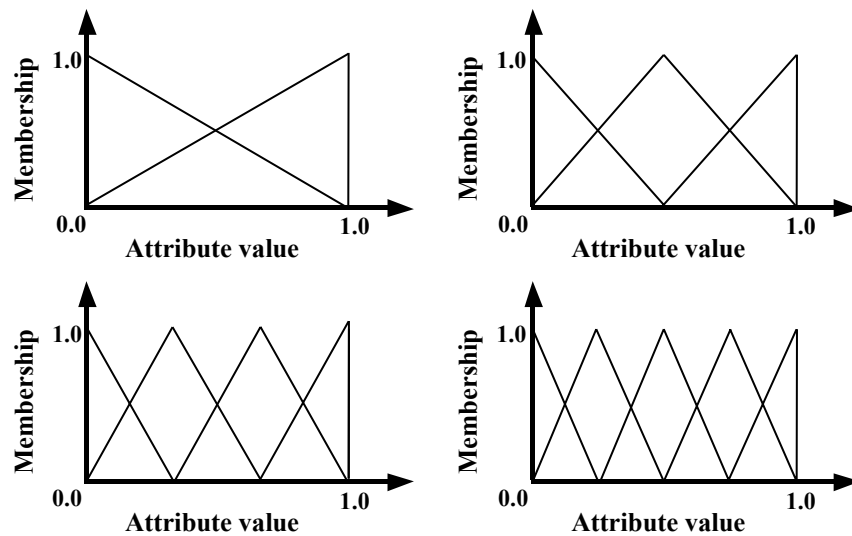
ファジィ相関ルール

Rule R_q : If x_{p1} is A_{q1} and ... and x_{p10} is A_{q10}
then y_1 is B_{q1} and ... and y_3 is B_{q3} .

- $\mathbf{x}_p = (x_{p1}, \dots, x_{p10})$: n 次元パターンベクトル
- R_q : q 番目のルールを示すラベル
- $\mathbf{A}_q = (A_{q1}, \dots, A_{q10})$: 条件部集合
- $\mathbf{B}_q = (B_{q1}, \dots, B_{q3})$: 結論部集合

30

条件部ファジィ集合



31

数値実験設定

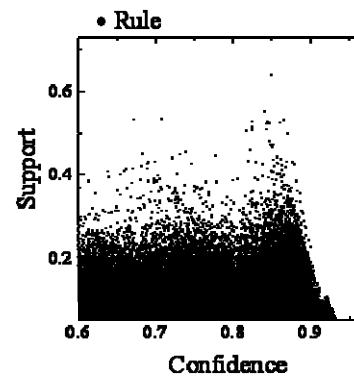
- Minimum confidence: 0.6
- Minimum support: 0.05
- 条件部最大ルール長: 3
- 結論部最大ルール長: 1

32

獲得されたルール

ルール数: 172033

時間(秒): 370.5



33

獲得されたルールの例

条件部の長さ1のルール

Supportの高い順

If 前桁高さ1 is *Small*² then 空力抵抗 is *Small*².

If 前桁高さ2 is *Small*² then 空力抵抗 is *Small*².

If XmaxTC1 is *Large*² then 空力抵抗 is *Small*².

If 前桁高さ2 is *Small*² then 衝撃波強さ is *Small*².

If 前桁高さ1 is *Small*² then 衝撃波強さ is *Small*².

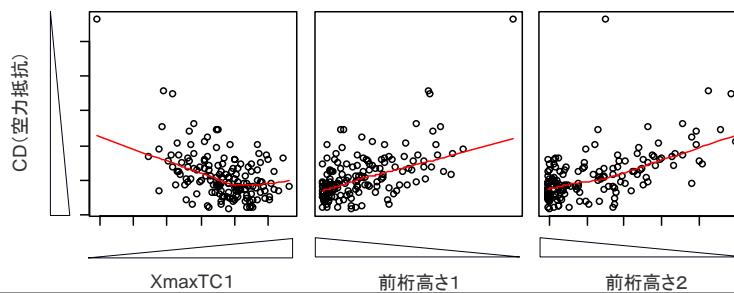
34

散布図との比較

If XmaxTC1 is *Large*² then 空力抵抗 is *Small*².

If 前桁高さ1 is *Small*² then 空力抵抗 is *Small*².

If 前桁高さ2 is *Small*² then 空力抵抗 is *Small*².

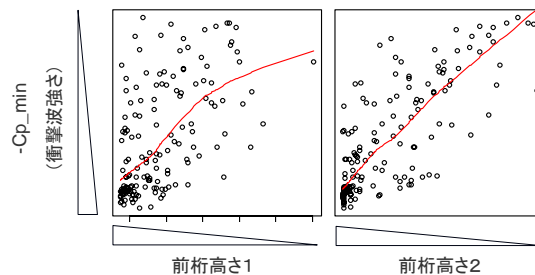


35

散布図との比較

If 前桁高さ1 is *Small*² then 衝撃波強さ is *Small*².

If 前桁高さ2 is *Small*² then 衝撃波強さ is *Small*².



36

獲得されたルール例

条件部の長さ1のルール

Confidenceの高い順

If XmaxTC2 is *Large*⁵ then 衝撃波強さ is *Small*².

If XmaxTC2 is *Large*⁵ then 空力抵抗 is *Small*².

If XmaxL2 is *Large*⁴ then 空力抵抗 is *Small*².

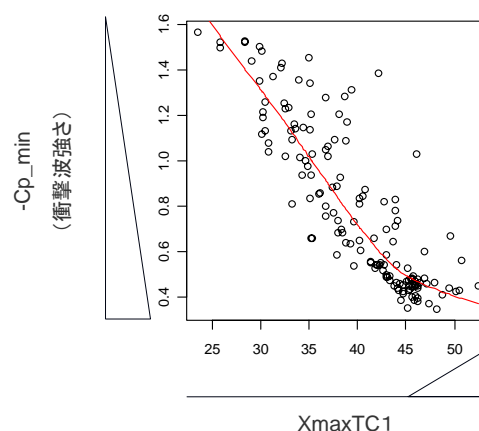
If XmaxL1 is *Large*⁵ then 空力抵抗 is *Small*².

If maxTC1 is *Small*⁵ then 空力抵抗 is *Small*².

37

散布図との比較

If XmaxTC2 is *Large*⁵ then 衝撃波強さ is *Small*².



38

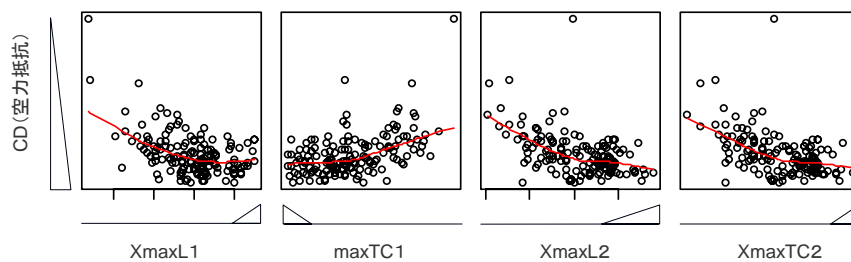
散布図との比較

If $X_{\max L1}$ is *Large*⁵ then 空力抵抗 is *Small*².

If $\max TC1$ is *Small*⁵ then 空力抵抗 is *Small*².

If $X_{\max L2}$ is *Large*⁴ then 空力抵抗 is *Small*².

If $X_{\max TC2}$ is *Large*⁵ then 空力抵抗 is *Small*².



39

他の評価基準

ConfidenceとSupportを単一で評価することは難しい.

他の評価基準

gain, variance, chi-squared value, entropy gain, gini, laplace, lift, conviction

パレート最適ルールは他の評価基準も最大化する.

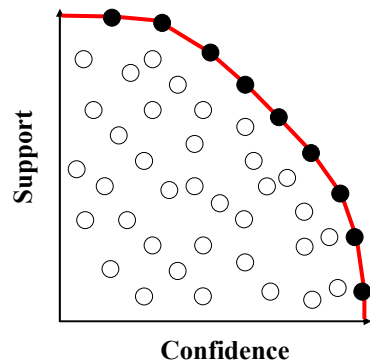
40

パレート最適ルール

定義

パレート最適ルール:

ConfidenceとSupportの値に関して他のルールに優越されないルール.



●: パレート最適ルール

41

パレート最適ルール

特徴

- 対象を絞り込むことができる.
- 他の評価値 (Lift値等) が高いルールが得られる.

Lift値

結論部が自明なルールは値が低くなる.

Amazonのお勧めシステム等でも用いられる.

42

パレート最適ルール

結果

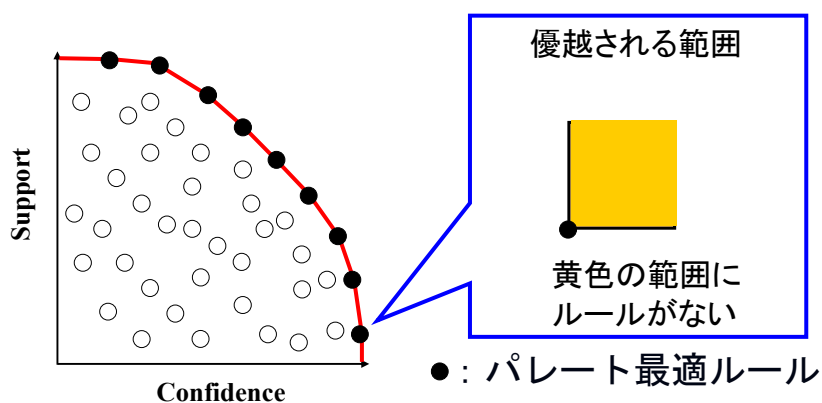
- 非常に数が少ない.
- 全てがSupport, Confidenceの値の高いルールに含まれていた.



パレート最適に近いルールも考慮する.

43

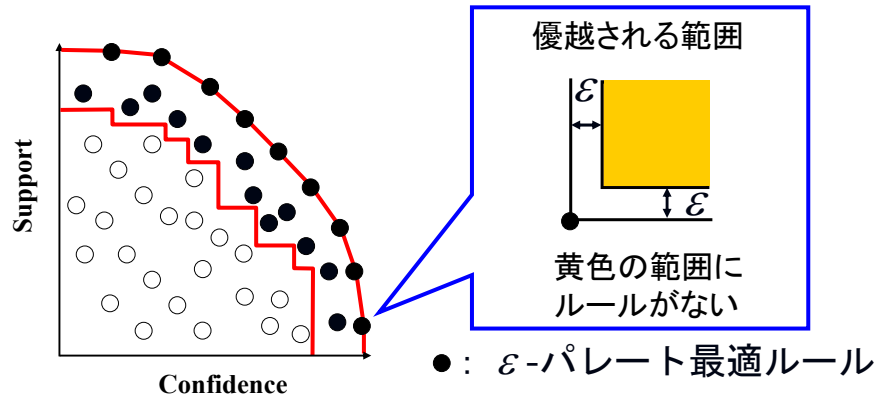
パレート最適ルール



黄色の範囲に他のルールが無いルール

44

ε - パレート最適ルール



黄色の範囲を右上に平行移動

45

ε - パレート最適ルール

5%だけ緩和した ε - パレート最適ルール(一部)

If XmaxL1 is *Large*⁵ then 衝撃波強さ is *Small*².
 If XmaxL2 is *Small*⁴ then 衝撃波強さ is *Large*².
 If maxTC1 is *Large*⁵ then 空力抵抗 is *Small*².
 If maxTC2 is *Large*⁵ then 空力抵抗 is *Small*².
 If maxL1 is *Medium*³ then 衝撃波強さ is *Small*².
 If XmaxTC1 is *Medium*³ then 衝撃波強さ is *Small*².

46

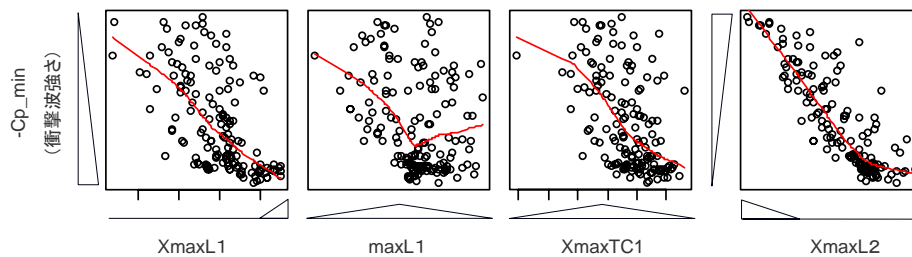
散布図との比較

If XmaxL1 is *Large*⁵ then 衝撃波強さは *Small*².

If maxL1 is *Medium*³ then 衝撃波強さは *Small*².

If XmaxTC1 is *Medium*³ then 衝撃波強さは *Small*².

If XmaxL2 is *Small*⁴ then 衝撃波強さは *Large*².

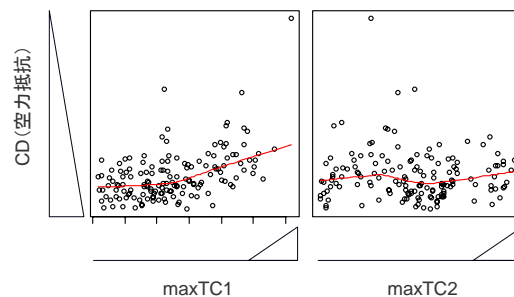


47

散布図との比較

If maxTC1 is *Large*⁵ then 空力抵抗 is *Small*².

If maxTC2 is *Large*⁵ then 空力抵抗 is *Small*².



48

獲得されたルール例

条件部の長さ2のルール

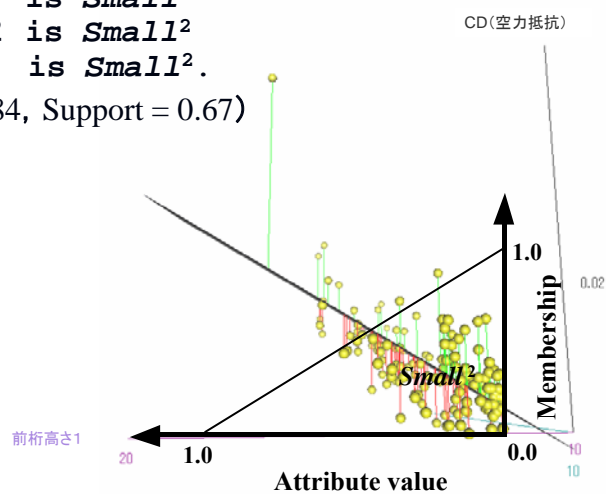
Supportの値の高いルール: Support = 0.53

If 前桁高さ1 is *Small*²
and 前桁高さ2 is *Small*²
then 空力抵抗 is *Small*².

49

散布図との比較

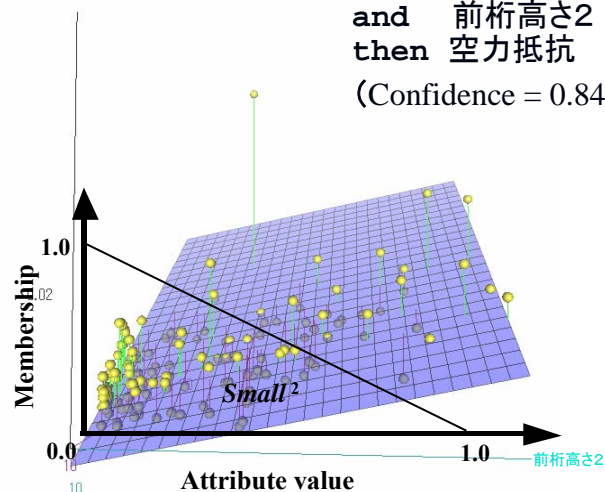
If 前桁高さ1 is *Small*²
and 前桁高さ2 is *Small*²
then 空力抵抗 is *Small*².
(Confidence = 0.84, Support = 0.67)



50

散布図との比較(別角度)

CD(空力抵抗)



If 前桁高さ1 is *Small*²
and 前桁高さ2 is *Small*²
then 空力抵抗 is *Small*².
(Confidence = 0.84, Support = 0.67)

51

獲得されたルール例

条件部の長さ2のルール

Confidenceの値の高いルール: Confidence = 0.93

If XmaxL1 is *Large*⁵
and maxL2 is *MediumSmall*⁵
then 衝撃波強さ is *Small*².

52

散布図との比較

-Cp_min(衝撃波強さ)

2

0.3

10

XmaxL1

If XmaxL1 is Large⁵
and maxL2 is Medium Small⁵
then 衝撃波強さ is Small².

(Confidence = 0.93, Support = 0.06)

この部分に対応

53

マイニング結果のまとめ

- ファジィルールにより知識が言語的に表現可能であることを示した.
- Supportの高いルールは一般的なルールであり, Confidenceの高いルールは特定の部分を正確に表すルールであることを示した.
- 翼の構造重量が結論部であるルールが獲得されにくかった.

54

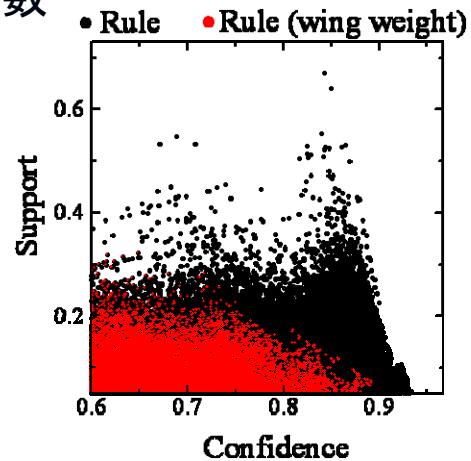
ルールの分布

結論部ごとのルール数

空力抵抗: 93697

衝撃波強さ: 50909

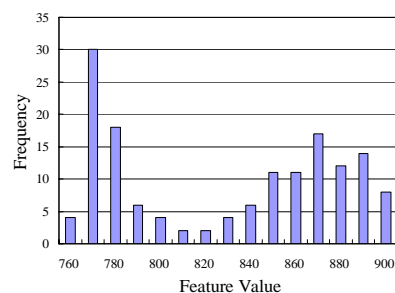
翼の重量: 27426



55

翼の構造重量

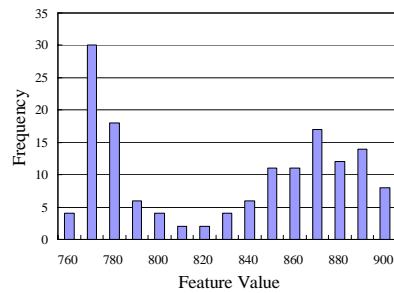
翼の構造重量についてさらに調査を行う.



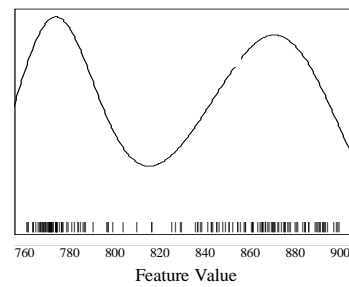
56

翼の構造重量

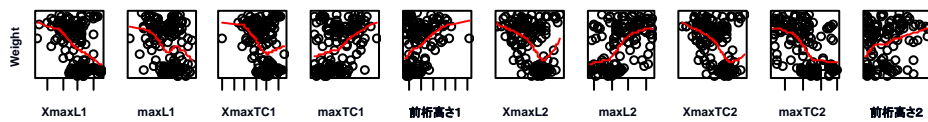
ヒストグラム



密度の分布



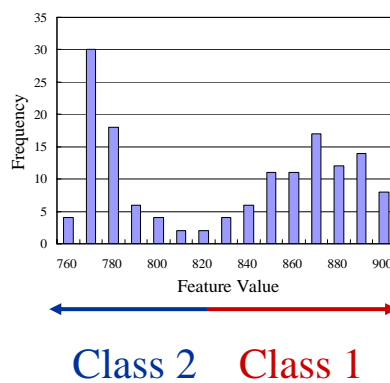
散布図



57

翼の構造重量

2つの山を分割してクラスとする.



58

○進化型多目的ルール選択

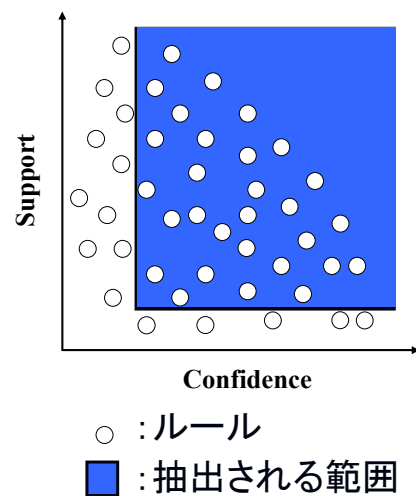
相関ルールマイニング手法

特徴

評価基準の閾値を満たすルールを全て抽出する.

評価基準:

- Confidence
- Support



60

相関ルールマイニング手法

問題点

- 非常に多くのルールが生成される.



計算コストの増加

- 識別器が複雑化する.

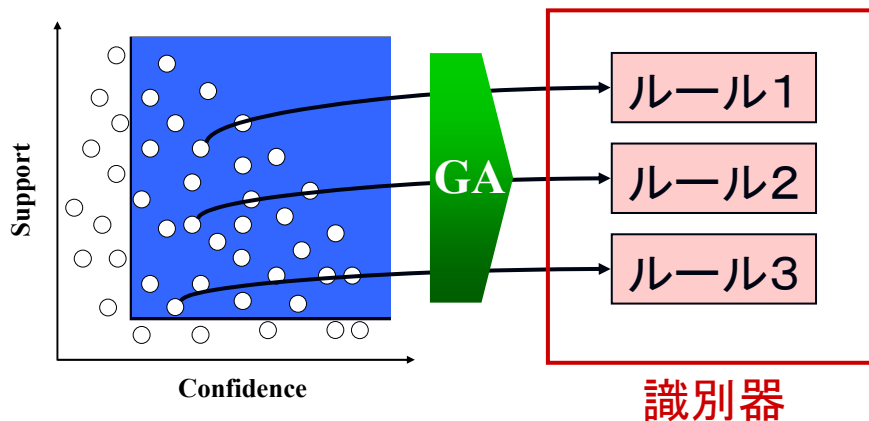


未知データに対する汎化能力の低下

61

進化型多目的ルール選択

抽出されたルールの中から, 遺伝的アルゴリズム (GA) を用いてルールを選択する.



62

進化型多目的ルール選択

特徴

- ルール数が少ない

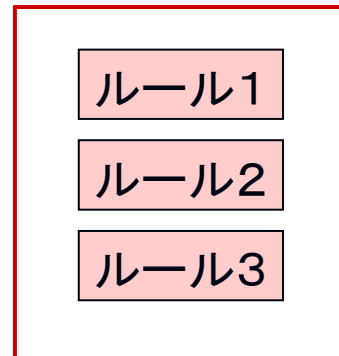


計算コストの減少

- 識別器が単純である



汎化能力の向上



識別器

63

識別ルール

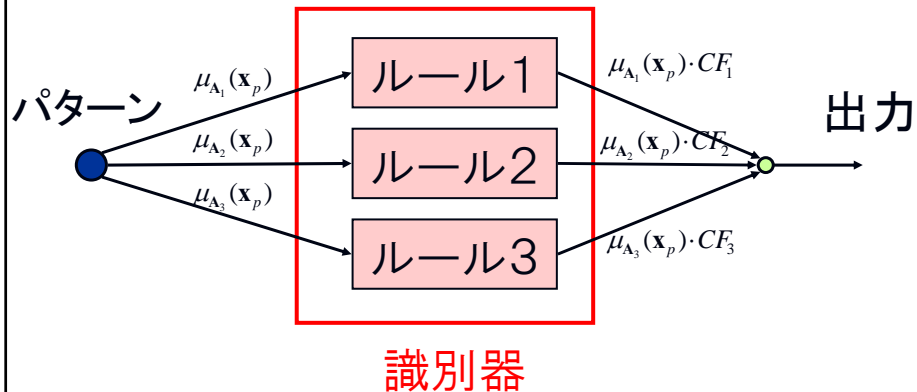
Rule R_q : If x_{p1} is A_{q1} and ... and x_{pn} is A_{qn}
then Class C_q with CF_q

- $\mathbf{x}_p = (x_{p1}, \dots, x_{pn})$: n 次元パターンベクトル
- R_q : q 番目のルールを示すラベル
- $\mathbf{A}_q = (A_{q1}, \dots, A_{qn})$: 条件部集合
- C_q : 結論部クラス
- CF_q : ルールの重み

64

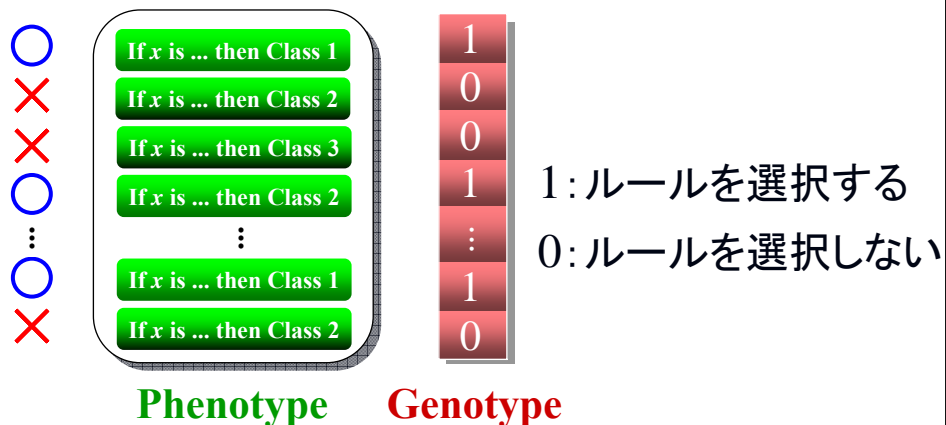
出力クラスの設定

メンバーシップ関数値と重みの積が最大のルール
の結論部を割り当てる。



65

ルール選択: コーディング



長さ N のバイナリ記号列でルールの選択を
表現する(N : 抽出されたルール数)。

66

目的関数

目的

識別器の精度を最大化し、複雑性を最小化する。

最大化：識別器 S の正答パターン数

最小化：識別器 S のルール数

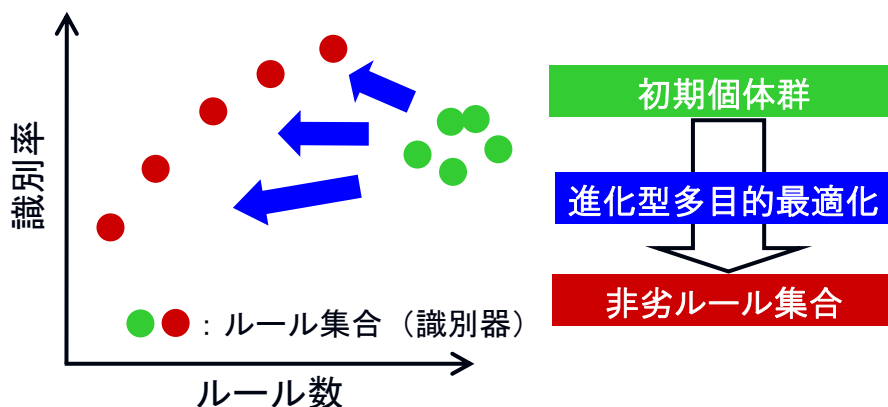
識別器 S の総ルール長

(ルール長：条件部ファジィ集合の総数)

67

進化型多目的最適化

進化型多目的最適化手法により、
3つの目的を同時に最適化する。



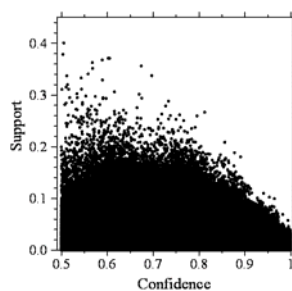
68

○ 数値実験

数値実験

クラス分布

- Class 1 83パターン (55.70%)
- Class 2 66パターン (44.30%)



70

パラメータ設定

進化型多目的最適化手法: **NSGA-II**

個体群サイズ: **200**

世代数: **1000** 世代

交叉確率: **0.9** (一様交叉)

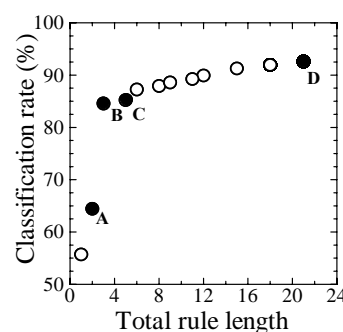
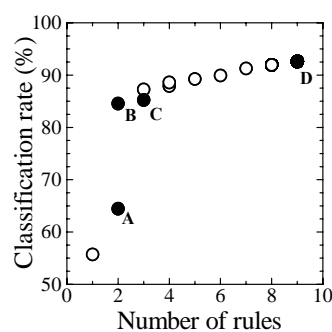
突然変異確率: **0.05** ($0 \rightarrow 1$), $1/N$ ($1 \rightarrow 0$)

N : 抽出されたルール数

ConfidenceとSupportの閾値の設定

Minimum Confidence: 0.6 Minimum Support: 0.01 71

獲得されたルール集合



ルール数と識別率





総ルール長と識別率

ルール集合A～Dを検証する.

72

ルール集合A

ルール集合の形式

	XmaxL1 maxL2	Consequent (Confidence, Support)
R_1	 	Class 1 (0.70, 0.34)
R_2	 	Class 2 (0.74, 0.09)

識別率:64.43%, ルール数:2, 総ルール長:2

R_1 : If maxL2 is *Large*² then Weight > 815.

R_2 : If XmaxL1 is *Large*⁵ then Weight < 815.

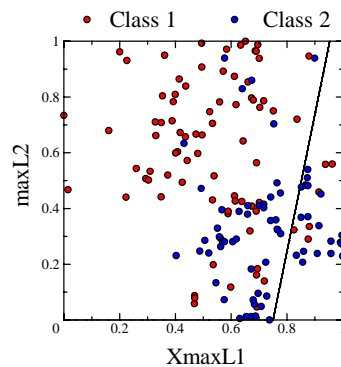
73

識別境界

R_1 : If maxL2 is *Large*² then Class 1.

R_2 : If XmaxL1 is *Large*⁵ then Class 2.

識別率:64.43%, ルール数:2, 総ルール長:2



精度はそれほど高くない。

74

ルール集合B

ルール集合の形式

	maxTC1	maxL2	XmaxTC2	Consequent (Confidence, Support)
R_1	DC		DC	Class 1 (0.70, 0.34)
R_2			DC	Class 2 (0.80, 0.18)

識別率:84.56%, ルール数:2, 総ルール長:3

R_1 : If maxL2 is *Large*² then Weight > 815.

R_2 : If maxTC1 is *Medium Large*⁵ and XmaxTC2 is *Medium Large*⁵ then Weight < 815.

75

ルール集合C

	maxTC1	maxL2	XmaxTC2	maxTC2	前桁高さ2	Consequent (Confidence, Support)
R_1	DC		DC	DC	DC	Class 1 (0.70, 0.34)
R_2			DC			Class 2 (0.80, 0.18)
R_3	DC	DC	DC			Class 2 (0.78, 0.05)

識別率:85.23%, ルール数:3, 総ルール長:5

76

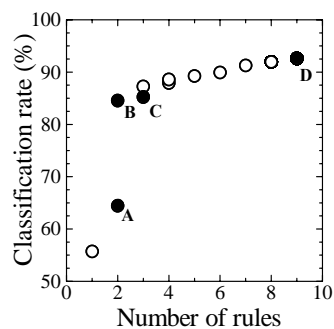
ルール集合D

	XmaxL1	XmaxTC1	maxTC1	前析高さ1	XmaxL2	maxL2	XmaxTC2	maxTC2	前析高さ2	Consequent (Confidence, Support)
R_1	▲	▲	▲	DC	DC	DC	DC	DC	DC	Class 2 (0.74, 0.06)
R_2	DC	DC	DC	DC	▲	DC	DC	DC	DC	Class 1 (0.80, 0.26)
R_3	DC	DC	DC	DC	DC	▲	DC	DC	DC	Class 1 (0.70, 0.34)
R_4	DC	DC	▲	▲	DC	▲	DC	DC	DC	Class 1 (0.92, 0.03)
R_5	DC	▲	DC	DC	DC	DC	▲	DC	DC	Class 2 (0.80, 0.18)
R_6	DC	DC	DC	▲	▲	DC	▲	DC	DC	Class 1 (0.88, 0.02)
R_7	▲	DC	DC	▲	DC	DC	DC	▲	DC	Class 2 (0.89, 0.03)
R_8	DC	DC	DC	DC	DC	DC	DC	▲	▲	Class 2 (0.78, 0.05)
R_9	DC	DC	DC	DC	DC	DC	▲	▲	▲	Class 2 (0.87, 0.15)

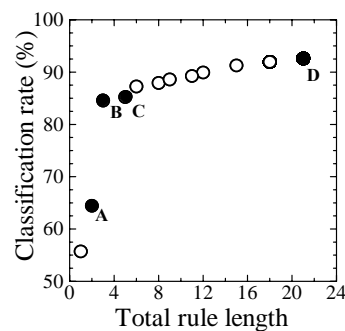
識別率: 92.62%, ルール数: 9, 総ルール長: 21

77

獲得されたルール集合



ルール数と識別率



総ルール長と識別率

進化型多目的ルール選択により, 利用者の要望に応じた知識獲得が可能になる.

78

○ おわりに

おわりに

- ファジィ相関ルールマイニングにより、多目的設計データからファジィルールの獲得を行った.
- ファジィ集合によりルールが言語的に表現されることを示した.
- Support, Confidence 等の評価基準に基づいてルールを示した.

おわりに

- パレート最適ルールの有用性を調査した.
- 進化型多目的ルール選択により, 利用者が望む知識を選択することができることを示した.

81

今後の課題

ファジィ相関ルールマイニング

- ルール数が多い.
- 埋もれているルールが存在する.
- 最大ルール長の制限がある.

進化型多目的ルール選択

- 教師なし学習への適用法の検討が必要.

82

今後の課題

- どのような知識が得られたのか？
- 自明な知識でないか？
- 設計に利用できる有用な知識か？

ConfidenceやSupport等の数値的指標だけでなく、利用者による評価が必要.

83

.....
Thank you
for your
kind attention.
.....

