身体動作の個人差に対してロバストな特徴量空間評価関数の提案

森 雅也[1], 秋月 拓磨[2], 高橋 弘毅[3], 大前 佑斗[1]

1: 東京工業高等専門学校 電気工学科

2: 豊橋技術科学大学

3:長岡技術科学大学大学院 工学研究科 情報・経営システム工学専攻

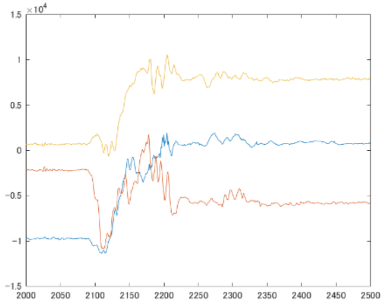
背景

慣性センサと機械学習による身体動作判定が盛ん

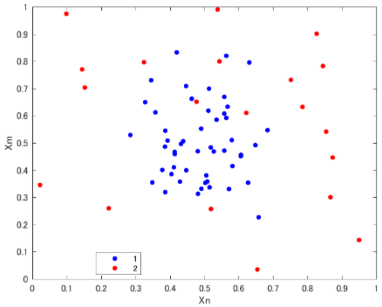
●3軸加速度センサを用いた人間行動認識[1]

●スイミングスタイルの自動分類[2]

身体動作判定に使用される慣性センサ



慣性センサから取得した身体動作の波形



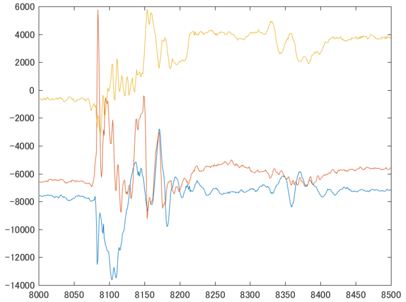
身体動作の波形から抽出された特徴量で構成された特徴量空間

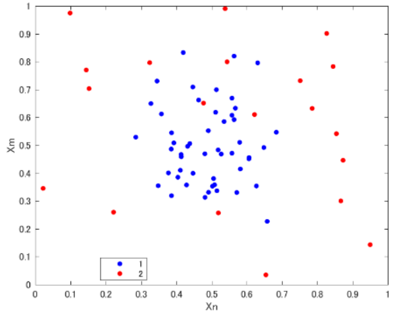
身体動作判定

[1] Khan, A. M., Lee, Y. K., Lee, S. Y., Kim, T. S., ``A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer'', IEEE transactions on information technology in biomedicine, 14(5), pp.1166-1172, 2010.
[2] Omae, Y., Kon, Y., Kobayashi, M., Sakai, K., Shionoya, A., Takahashi, H., Akiduki, T., Nakai, K., Ezaki, N., Sakurai, Y., Miyaji, C., ``Swimming Style Classification Based on Ensemble Learning and Adaptive Feature Value by Using Inertial Measurement Unit'', Journal of Advanced Computational Intelligence and Intelligent Informatics, 21(4), pp.616-631, 2017.

特徴量探索の問題点

- 身体動作には個人差が含まれている
- 既存の特徴量評価関数には個人差が考慮されていない
- 多くの人に対して良い特徴量を選出できるとは限らない





Very good

既存手法による最適な特徴量探索

- クラス内分散・クラス間分散比[3]
- MRS[4]
- ・・・

1番良い空間
(Very good)
と評価

しかし
個別に確認すると…

既存手法

被験者1：Very good
被験者2：Bad
被験者3：Bad
被験者4：Very good
⋮

全員に良いとは限らない

[3] 荒木雅弘, “フリーソフトでつくる音声認識システム: パターン 認識・機械学習の初歩から対話システムまで”, 森北出版, 2007.
[4] Chen, X. W., Jong, C. J., “Minimum Reference Set Based Feature Selection for Small Sample Classifications”, The 24th International Conference on Machine Learning, pp.153-160, 2007.

目的

特定個人のみならず，**多くの人に対して良いといえるような特徴量空間を良いと判断できる**特徴量空間評価関数の提案

既存手法

被験者1：**Very Good**
被験者2：Bad
被験者3：Bad
被験者4：**Very Good**
・
・
・

良し悪しが混合した結果を返すような特徴量選択

提案手法

被験者1：Good
被験者2：Good
被験者3：Good
被験者4：Good
・
・
・

多くの人に対し、それなりに良い結果を返すような特徴量選択

提案手法の適用範囲

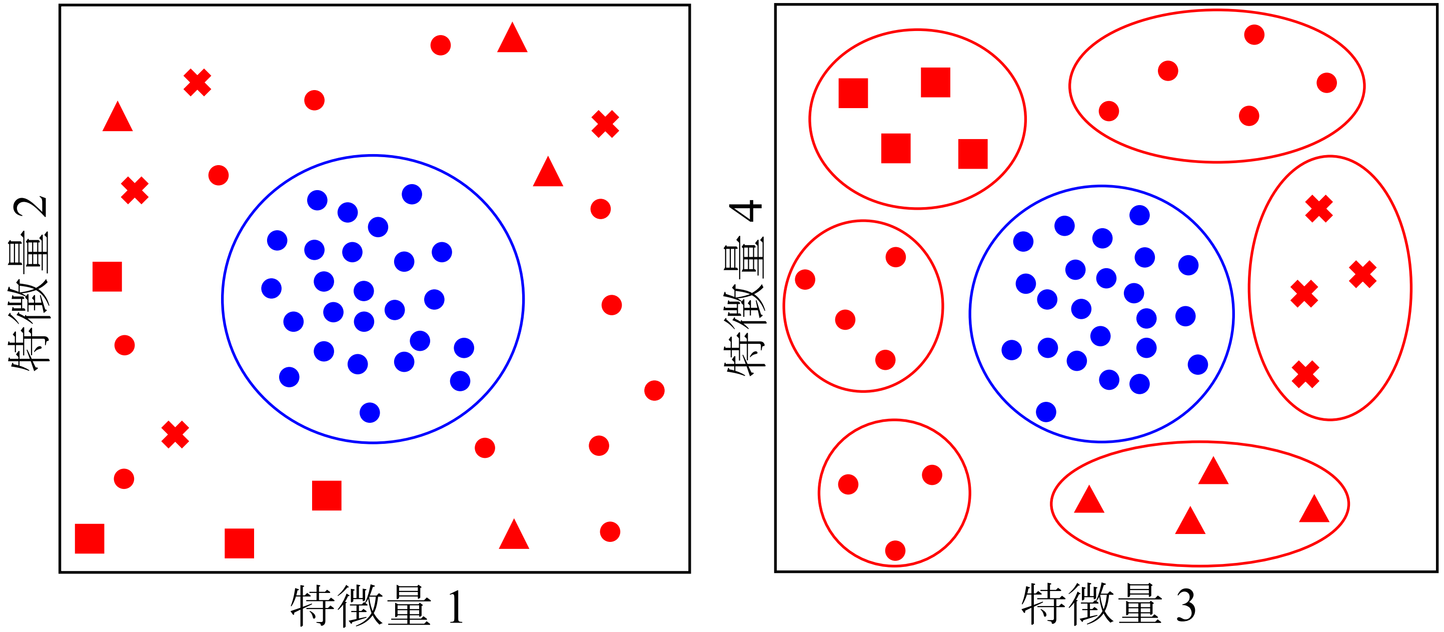
異常値検知問題(a):
検出したい身体動作を正常、それ以外を異常とし、いずれかを自動判定する問題

クラス分類問題(b):
様々な身体動作を個別のクラスとみなし、それらを判定する問題

本研究では、**異常値検知問題**において多くの被験者に優れた空間を探索するための手法を提案する。

(a)
異常値検知問題として
優れた特徴量空間

(b)
クラス分類問題として
優れた特徴量空間



凡 例

●: 正常な運転 (正常)
✖: わき見運転 (異常1)
▲: 片手運転 (異常2)
■: スマートフォンの利用 (異常3)
●: その他さまざまな異常運転 (異常n)

考えている問題

$F_{\max}$個の特徴量$\{x_1, x_2, \cdots, x_{F_{\max}}\}$があり、
N_{sub} 人の被験者$\{1, 2, \cdots, N_{\mathrm{sub}}\}$から
正常(nor),異常(ano)データが取れているとする。
このとき、全員に重要な2本の特徴量 $< x_1^{\mathrm{opt}}, x_2^{\mathrm{opt}} >$を探す
最適化問題を考える。
解のパターン数は${}_{F_{\max}}C_2$ 通りある。

多次元に拡張できるように構成

ここで、

$$x_n, x_m \in X := \{x_1, x_2, \cdots, x_{F_{\max}}\}$$

$$i \in I := \{1, 2, \cdots, N_{\mathrm{sub}}\}$$

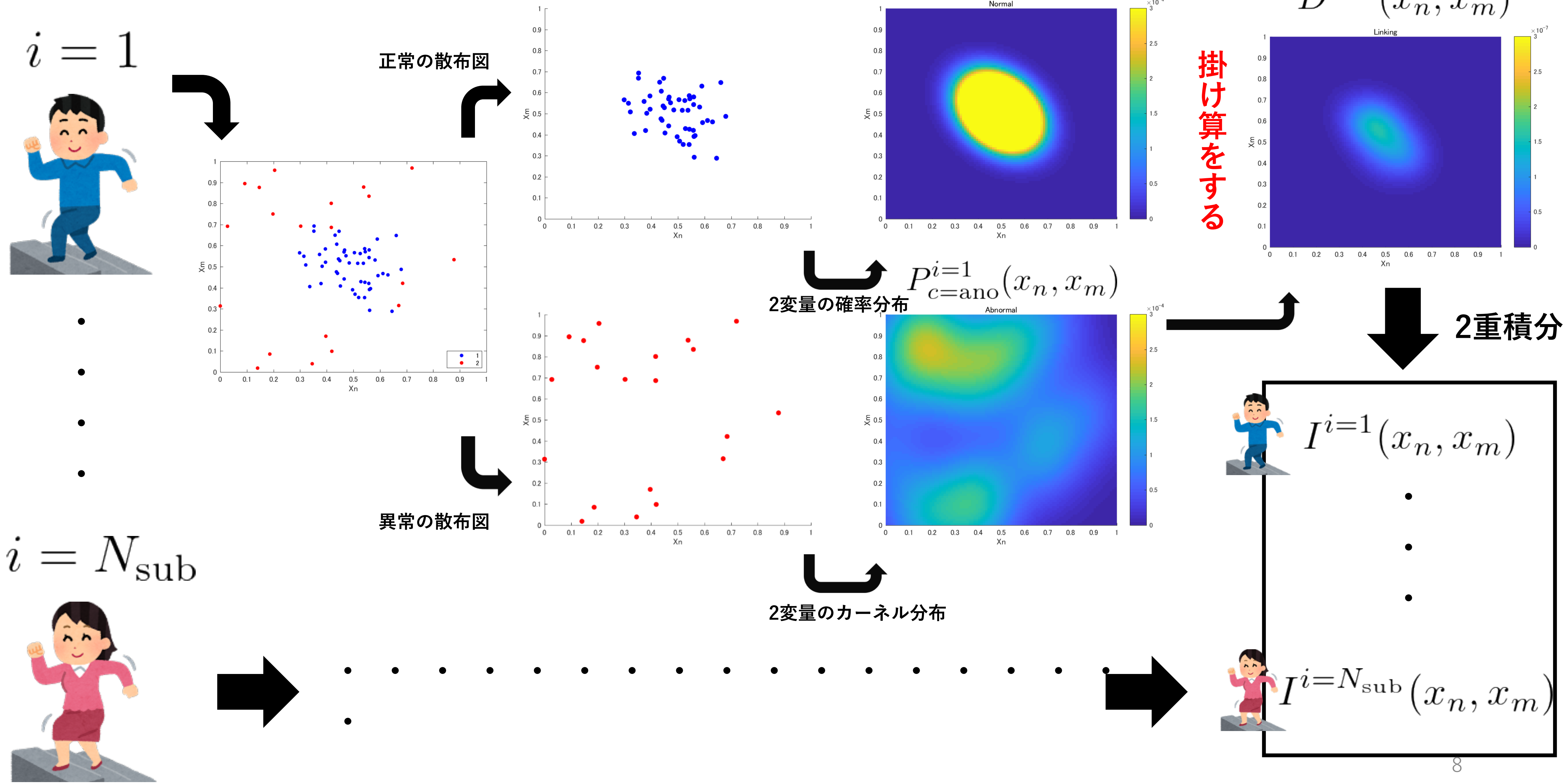
$$c \in C := \{\mathrm{ano}, \mathrm{nor}\}$$

とする。

提案手法:

1. ある特徴量空間における1人の被験者の身体動作データ（正常・異常）をプロット
2. 正常・異常で、プロットを2つの空間に分割
3. 正常のプロットには2変量正規分布を、異常のプロットには2変量カーネル分布を適用し、確率密度関数を導出
4. 正常と異常の確率密度関数を掛け合わせ、重複関数を導出
5. 重複関数を2重積分し、誤分類危険度(1次元の実数)を導出

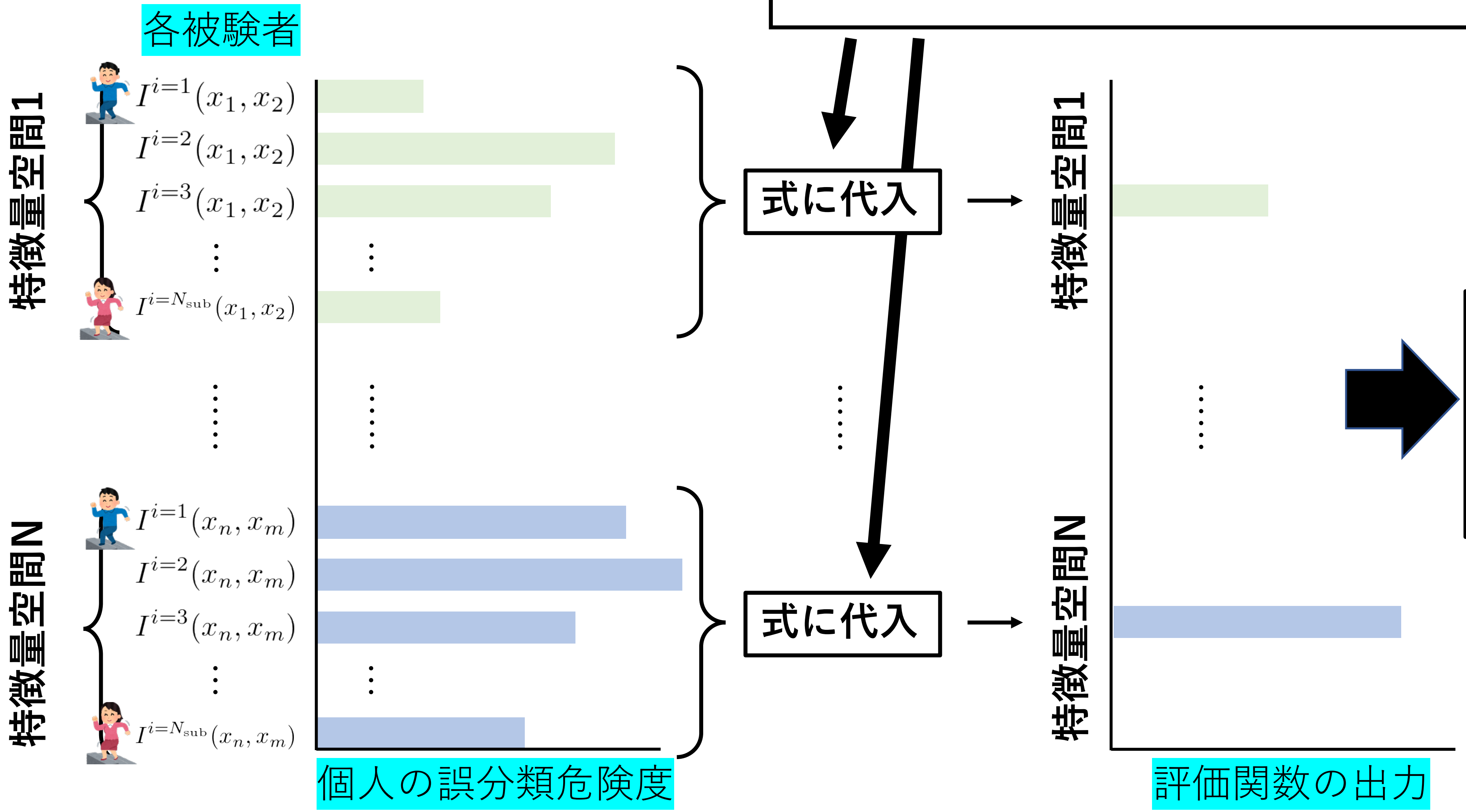
被験者 i の$< x_n, x_m >$の誤分類危険度



提案手法:

$$\alpha \ \mathrm{mean}\{I(x_u, x_v)\}' + \beta \ \mathrm{std}\{I(x_u, x_v)\}'$$

$\underbrace{\alpha \ \mathrm{mean}\{I(x_u, x_v)\}'}_{\text{全員の誤分類危険度の平均}} + \beta \ \underbrace{\mathrm{std}\{I(x_u, x_v)\}'}_{\text{全員の誤分類危険度の標準偏差}}$

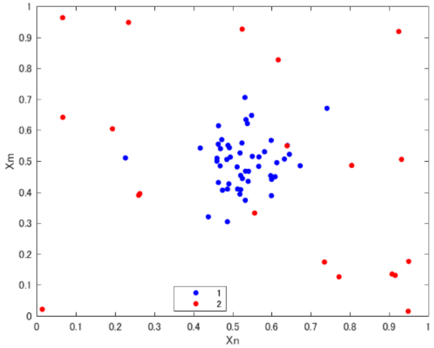


評価関数の出力が<u>**1番小さい**</u>時の特徴量空間を最適とする

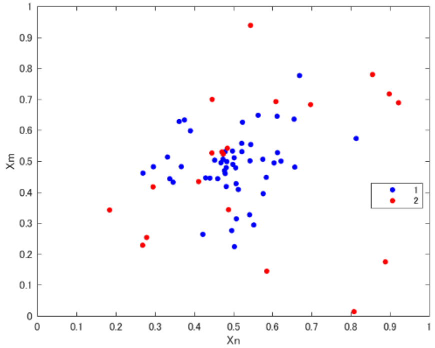
実験概要:

提案手法の妥当性を検証するため、4パターンのダミーデータを生成

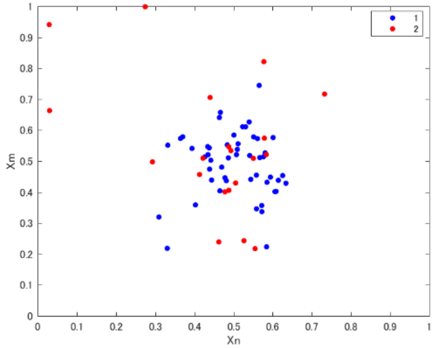
Very Good



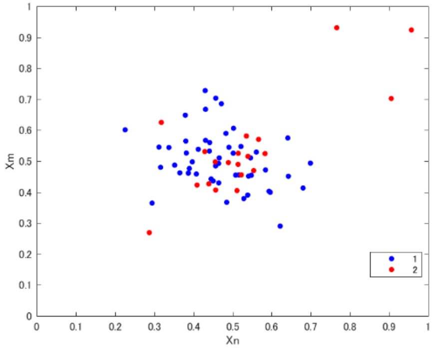
Good



Bad



Very Bad



	正常データ (50 plot)	異常データ (20 plot)
Very Good	Mean=(0.5, 0.5), Std=(0.1, 0.1), Cov=(0, 0)の2変量正規分布に従う乱数	Mean ± 1 Std 以内に0プロット存在する一様乱数
Good		Mean ± 1 Std 以内に5プロット存在する一様乱数
Bad		Mean ± 1 Std 以内に10プロット存在する一様乱数
Very Bad		Mean ± 1 Std 以内に15プロット存在する一様乱数

生成データのパターン

Case	Subjects					提案手法 (α, β)= (0.7,0.3)	提案手法 (α,β)= (0.5,0.5)	クラス内分散 クラス間分散比	MRS
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5				
01	**Very Good**	**Very Good**	**Very Good**	**Very Good**	**Very Good**	**1**	**1**	**3**	**1**
02	Good	Good	Good	Good	Good	**3**	**2**	11	6
03	Bad	Bad	Bad	Bad	Bad	7	6	8	14
04	**Very Bad**	**Very Bad**	**Very Bad**	**Very Bad**	**Very Bad**	12	9	12	7
05	**Very Good**	**Very Good**	**Very Good**	Good	Good	**2**	**3**	9	**2**
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Bad	**4**	7	**2**	**4**
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**Very Bad**	10	12	10	**5**
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**Very Good**	11	13	14	13
09	**Very Bad**	**Very Bad**	**Very Bad**	Good	Good	14	14	**1**	11
10	**Very Bad**	**Very Bad**	**Very Bad**	Bad	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11	Good	Good	Good	Bad	Bad	**3**	**4**	13	**3**
12	Bad	Bad	Bad	Good	Good	6	**5**	6	10
13	**Very Good**	Good	Bad	**Very Bad**	Good	9	10	**5**	8
14	**Very Good**	Good	Bad	**Very Bad**	Bad	8	8	**4**	12

被験者2に対して、case02の特徴量空間(2次元)の異常・正常データの分布が『Good』であったことを意味する。

MRSの場合、case01の特徴量空間(2次元)が1番良いと評価されたことを意味する。

Case	Subjects					提案手法		既存手法	
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	$(\alpha,\beta)=(0.7,0.3)$	$(\alpha,\beta)=(0.5,0.5)$	クラス内分散 クラス間分散比	MRS
01	Very Good	Very Good	Very Good	Very Good	Very Good	**1**	**1**	**3**	**1**
02	Good	Good	Good	Good	Good	**3**	**2**	11	6
03	Bad	Bad	Bad	Bad	Bad	7	6	8	14
04	Very Bad	Very Bad	Very Bad	Very Bad	Very Bad	12	9	12	7
05	Very Good	Very Good	Very Good	Good	Good	**2**	**3**	9	**2**
06	Very Good	Very Good	Very Good	Bad	Bad	**4**	7	**2**	**4**
07	Vary Good	Vary Good	Vary Good	Very Bad	Very Bad	10	12	10	**5**
08	Very Bad	Very Bad	Very Bad	Very Good	Very Good	11	13	14	13
09	Very Bad	Very Bad	Very Bad	Good	Good	14	14	**1**	11
10	Very Bad	Very Bad	Very Bad	Bad	Bad	13	11	7	9
11	Good	Good	Good	Bad	Bad	**5**	**4**	13	**3**
12	Bad	Bad	Bad	Good	Good	6	**5**	6	10
13	Very Good	Good	Bad	Very Bad	Good	9	10	**5**	8
14	Very Good	Good	Bad	Very Bad	Bad	8	8	**4**	12

MRSの結果(既存手法):

- case01とcase05は望む結果が得られている
- 3番目に良いと思われるcase2が6位と評価されている

Case	Subjects					提案手法		既存手法	
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	(α,β)=(0.7,0.3)	(α,β)=(0.5,0.5)	クラス内分散クラス間分散比	MRS
01	Very Good	Very Good	Very Good	Very Good	Very Good	1	1	3	1
02	Good	Good	Good	Good	Good	3	2	11	6
03	Bad	Bad	Bad	Bad	Bad	7	6	8	14
04	Very Bad	Very Bad	Very Bad	Very Bad	Very Bad	12	9	12	7
05	Very Good	Very Good	Very Good	Good	Good	2	3	9	2
06	Very Good	Very Good	Very Good	Bad	Bad	4	7	2	4
07	Vary Good	Vary Good	Vary Good	Very Bad	Very Bad	10	12	10	5
08	Very Bad	Very Bad	Very Bad	Very Good	Very Good	11	13	14	13
09	Very Bad	Very Bad	Very Bad	Good	Good	14	14	1	11
10	Very Bad	Very Bad	Very Bad	Bad	Bad	13	11	7	9
11	Good	Good	Good	Bad	Bad	5	4	13	3
12	Bad	Bad	Bad	Good	Good	6	5	6	10
13	Very Good	Good	Bad	Very Bad	Good	9	10	5	8
14	Very Good	Good	Bad	Very Bad	Bad	8	8	4	12

クラス内分散クラス間分散比の結果(既存手法):

- クラス分類問題に使用される手法
- 異常値検知問題では望む結果を得ることはできない

Case	Subjects					提案手法		既存手法	
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	（α，β）=（0.7,0.3）	（α，β）=（0.5,0.5）	クラス内分散クラス間分散比	MRS
01	**Very Good**	**Very Good**	**Very Good**	**Very Good**	**Very Good**	**1**	**1**	**3**	**1**
02	Good	Good	Good	Good	Good	**3**	**2**	11	6
03	Bad	Bad	Bad	Bad	Bad	7	6	8	14
04	**Very Bad**	**Very Bad**	**Very Bad**	**Very Bad**	**Very Bad**	12	9	12	7
05	**Very Good**	**Very Good**	**Very Good**	Good	Good	**2**	**3**	9	**2**
06	**Very Good**	**Very Good**	**Very Good**	Bad	Bad	**4**	7	**2**	**4**
07	**Vary Good**	**Vary Good**	**Vary Good**	**Very Bad**	**Very Bad**	10	12	10	**5**
08	**Very Bad**	**Very Bad**	**Very Bad**	**Very Good**	**Very Good**	11	13	14	13
09	**Very Bad**	**Very Bad**	**Very Bad**	Good	Good	14	14	**1**	11
10	**Very Bad**	**Very Bad**	**Very Bad**	Bad	Bad	13	11	7	9
11	Good	Good	Good	Bad	Bad	**5**	**4**	13	**3**
12	Bad	Bad	Bad	Good	Good	6	**5**	6	10
13	**Very Good**	Good	Bad	**Very Bad**	Good	9	10	**5**	8
14	**Very Good**	Good	Bad	**Very Bad**	Bad	8	8	**4**	12

提案手法(α,β)=(0.5,0.5)の結果:

- α(平均の重み)とβ(標準偏差の重み)を同等にして検証
- 考察として、標準偏差が重要視されすぎた

Case	Subjects					提案手法		既存手法	
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	(α,β)=(0.7,0.3)	(α,β)=(0.5,0.5)	クラス内分散クラス間分散比	MRS
01	**Very Good**	**Very Good**	**Very Good**	**Very Good**	**Very Good**	**1**	**1**	**3**	**1**
02	Good	Good	Good	Good	Good	**3**	**2**	11	6
03	Bad	Bad	Bad	Bad	Bad	7	6	8	14
04	**Very Bad**	**Very Bad**	**Very Bad**	**Very Bad**	**Very Bad**	12	9	12	7
05	**Very Good**	**Very Good**	**Very Good**	Good	Good	**2**	**3**	9	**2**
06	**Very Good**	**Very Good**	**Very Good**	Bad	Bad	**4**	7	**2**	**4**
07	**Vary Good**	**Vary Good**	**Vary Good**	**Very Bad**	**Very Bad**	10	12	10	**5**
08	**Very Bad**	**Very Bad**	**Very Bad**	**Very Good**	**Very Good**	11	13	14	13
09	**Very Bad**	**Very Bad**	**Very Bad**	Good	Good	14	14	**1**	11
10	**Very Bad**	**Very Bad**	**Very Bad**	Bad	Bad	13	11	7	9
11	Good	Good	Good	Bad	Bad	**5**	**4**	13	**3**
12	Bad	Bad	Bad	Good	Good	6	**5**	6	10
13	**Very Good**	Good	Bad	**Very Bad**	Good	9	10	**5**	8
14	**Very Good**	Good	Bad	**Very Bad**	Bad	8	8	**4**	12

提案手法(α,β)=(0.7,0.3)の結果:

- 1位から5位まで望む結果が得られた
- case07よりcase11を重要視したことから、多くの人に対してそれなりに良い結果を返すような特徴量選択ができているということが確認できた

まとめ

発表内容：
特定個人のみならず，多くの人に対して良いといえるような特徴量空間を良いと判断できる特徴量空間評価関数を考案した。

提案手法の手続き：

[手続き1]　ある特徴量空間における**1人の**正常・異常データのプロットから、2つの**確率密度関数**を求める
[手続き2]　正常・異常の確率密度関数を**掛け合わせて、重複関数**を求める
[手続き3]　重複関数を2重積分して、ある特徴量空間における**1人の誤分類危険度**を求める
[手続き4]　**特徴量空間毎**に、**全被験者の誤分類危険度**を考案した評価関数に代入し、**1次元の実数**を求める
[手続き5]　求めた1次元の実数を順位付けし、**1番小さかった時**の特徴量空間を最適な空間だと定義する

適用結果と今後の展望：

- ダミーデータでの検証で、多くの人に対し、平均的に良い特徴量空間を選択していることが確認できた。
- ダミーデータでは望む結果を得ることができた

⟶実データに適用し、妥当性を検証する