

信号処理における 機械学習的クラスタリング手法の開発

情報工学基盤講座 麻生到

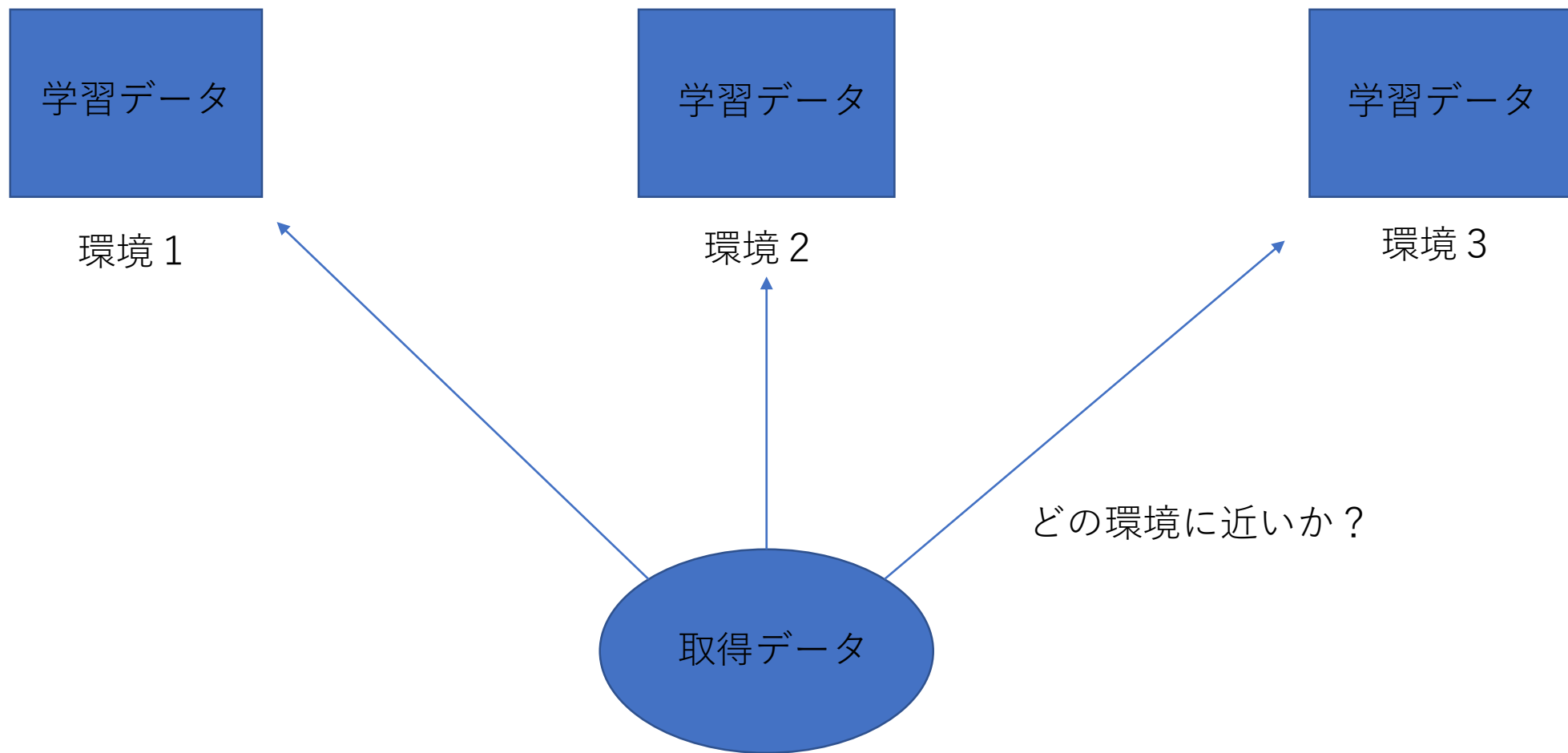
はじめに

センサや計測器から取得できる信号は、環境の変化の影響で常に同じような信号が取得できるとは限らない



環境の変化に対応できる処理が必要となる

適用手法



テストデータだったら、どの環境の学習データを使うか？

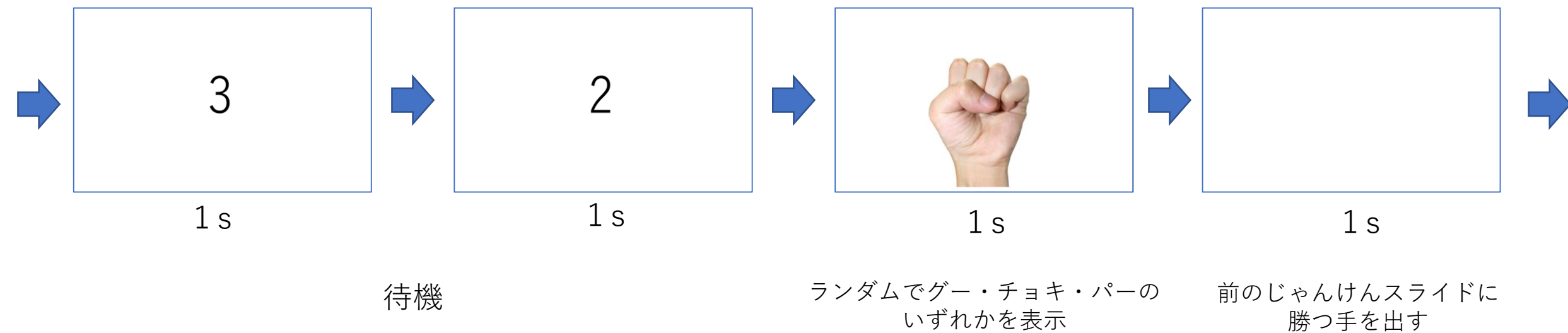
本研究の目的

環境の変化に対応できる信号処理システムの作成

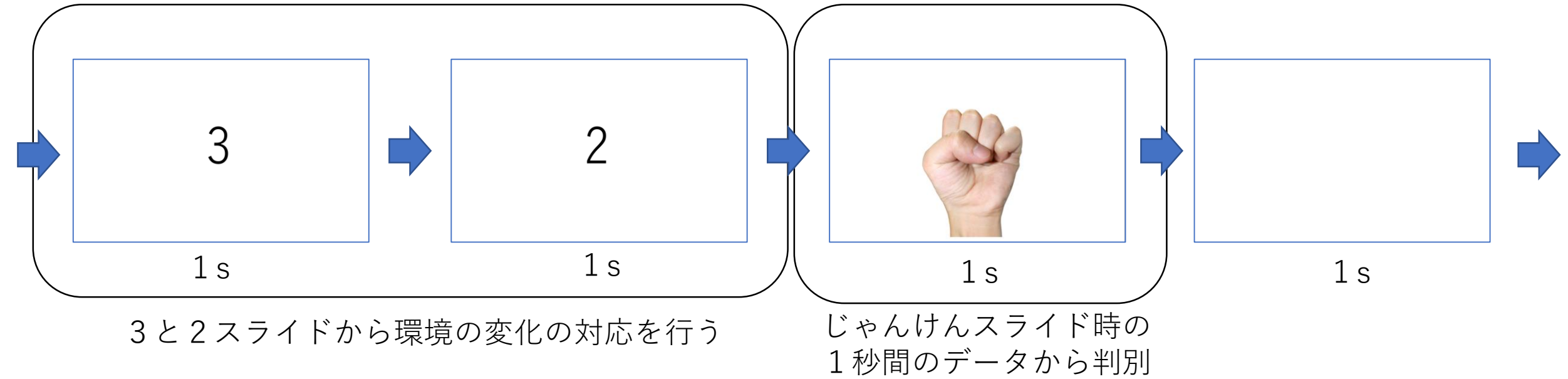


脳波データを用いてじゃんけんで使用者に勝つシステムの作成

じゃんけんシステムの流れ



スライド別の処理



クラスタリング

ベイズ理論を用いる

$$\text{事後分布} = \text{尤度関数} + \text{事前分布}$$

- ・尤度関数

あるパラメータが与えられた時に、データがどのような分布をしているか

- ・事前分布

どんなパラメータが与えられやすいか（パラメータの散らばり具合）

データの事後分布のパラメータがどのような値になるかを推定



事後分布を使うことで、未知のデータが予測可能

パラメータを推定することが目的

本研究では、高次元ベクトルが従う確率分布を学習

目的：脳波データから3クラス分類を行いたい（ゲー・チョキ・パー）

前処理

脳波データを3-50Hzのパワースペクトルに周波数解析



分類に用いるデータはパワースペクトル(94次元)

取得したパワースペクトルはガウス分布に従うと仮定

ガウス分布のパラメータ

- ・ 平均ベクトル
- ・ 分散共分散行列

つまり、3クラスそれぞれのガウス分布のパラメータを求め、確率分布を推定する

環境の判別

KLダイバージェンスを用いて2つの確率分布の類似度を求める



テストデータが学習データのどの環境に近いか

近い環境の学習データを用いることで分類精度が環境変動がない場合の精度と変わらないことが示すことができたなら、環境変動への対処ができていることを示せる

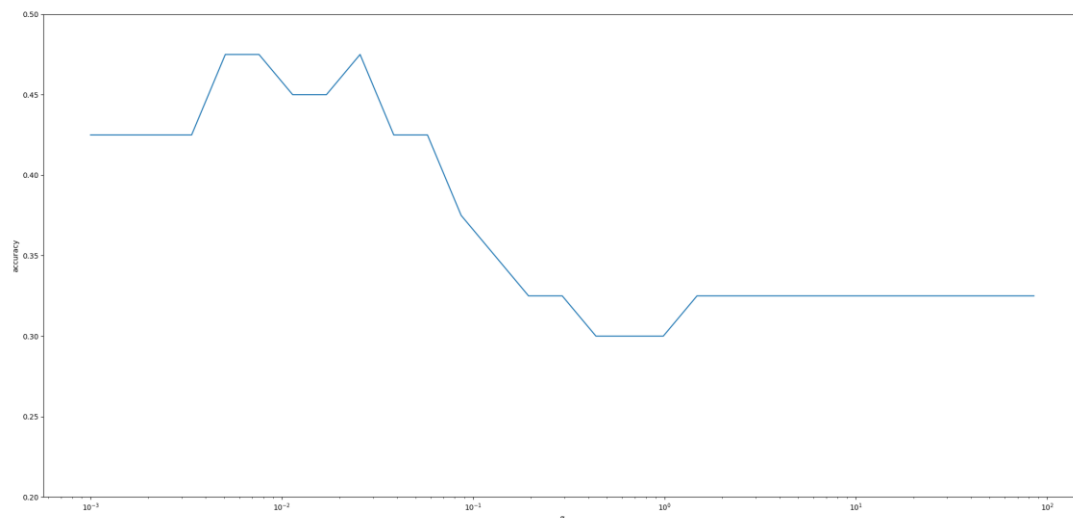
評価実験

環境変動なし

1 日分の環境変動なしの訓練データ60セットとテストデータ40セットを用いて分類を行った

$$W_0 = \alpha E$$

パラメータ α の値と精度の関係を以下のグラフに示す



$\alpha = 0.007$ あたりで分類精度が47.5%となっている

環境変動なし

環境の影響を考えない場合の4日間の学習データとテストデータの分類精度

	Day1	Day2	Day3	Day4
α	0.2919	0.0076	0.038	0.0051
train(accuracy)	0.517	0.367	0.517	0.418
test(accuracy)	0.450	0.475	0.450	0.475

学習データとテストデータとの精度の差があまり大きくない



過学習をしておらず、汎化性能のある分類器であることがわかる

環境変動あり

まず、KLダイバージェンスを用いた環境の選定

表のKLダイバージェンスの値は他3つの環境の中で最も小さい値

	Day1	Day2	Day3	Day4
α	0.2919	0.0076	0.038	0.0051
KL ダイバージェンス	11.5423	9.7056	10.3993	17.0014
test(環境変動への対応無し)	0.4083	0.3833	0.4083	0.3917
test(環境変動への対応有り)	0.425	0.425	0.425	0.375

環境変動への対応無しときより、環境変動への対応有りの場合のほうがDay1、Day2、Day3が精度が高い



環境の影響に対して、対応できていることが示唆される

環境変動あり

	Day1	Day2	Day3	Day4
α	0.2919	0.0076	0.038	0.0051
KL ダイバージェンス	11.5423	9.7056	10.3993	17.0014
test(環境変動への対応無し)	0.4083	0.3833	0.4083	0.3917
test(環境変動への対応有り)	0.425	0.425	0.425	0.375

精度が下がっている

KLダイバージェンスの値が大きいことから、適応できる学習データを用意できていない

適応できる学習データが存在すれば、精度が上がると考えれる

今後の課題

- ・デモの完成