

これまでの月震分類は正しいのか？ ～機械学習の視点から見た妥当性～

加藤広大(首都大学東京システムデザイン学部4年)
山田竜平(国立天文台RISE月惑星探査検討室)
山本幸生(宇宙航空研究開発機構宇宙科学研究所)
横山昌平(静岡大学情報学部)
石川博 (首都大学東京システムデザイン学部)



目次

🌙 背景

🌙 従来の震源分類の妥当性の検証

🌙 サブクラスの有無の検証

🌙 まとめ



目次

🌙 背景

🌙 従来の震源分類の妥当性の検証

🌙 サブクラスの有無の検証

🌙 まとめ



これまでの月震研究

🌙 月にも地震があった！

- 🌐 NASAのアポロミッションで月に地震計を設置
- 🌐 約7年半データ取得（1969年～1977年）



なぜ月震を解析するのか

月の内部構造がわかる

- 🌐 月震の発生メカニズム
- 🌐 震源
- 🌐 月震波形の特徴

月震には大きく4種類

深発月震
熱月震

浅発月震
隕石衝突

半数以上が深発月震
ノイズが多い
揺れが1時間以上続くものも



深発月震について

深発月震の特徴

- ☾ 周期的に発生
- ☾ 同一震源の月震は波形が類似



波形の類似性を考慮した
震源ラベルの付与

[Nakamura et al.,2003]
これが従来の震源分類の基準

従来の分類基準への疑問

- ☾ 従来とは異なる手法で月震の検出・分類に成功
 - 🌐 [Bulow et al.,2005], [Endrun et al.,2015]
- ☾ SOMを用いた月震の可視化システムで示唆
 - 🌐 [Goto et al.,2014]



本研究の目的

従来の分類基準への疑問

- 機械学習を用いた従来の月震分類の検証
- 震源ごとの特徴の分析

本研究の目的

大きな目標

- 従来の分類手法の見直し
- 分類基準の検討, 再定義

機械学習とは

データから反復的に分類器を訓練し、
そこに潜むパターンを見つけ出すこと

応用例

- スпам検知
- 文字認識
- 顔認証
- 商品レコメンデーション
- 株価予測
- 音声認識 etc.

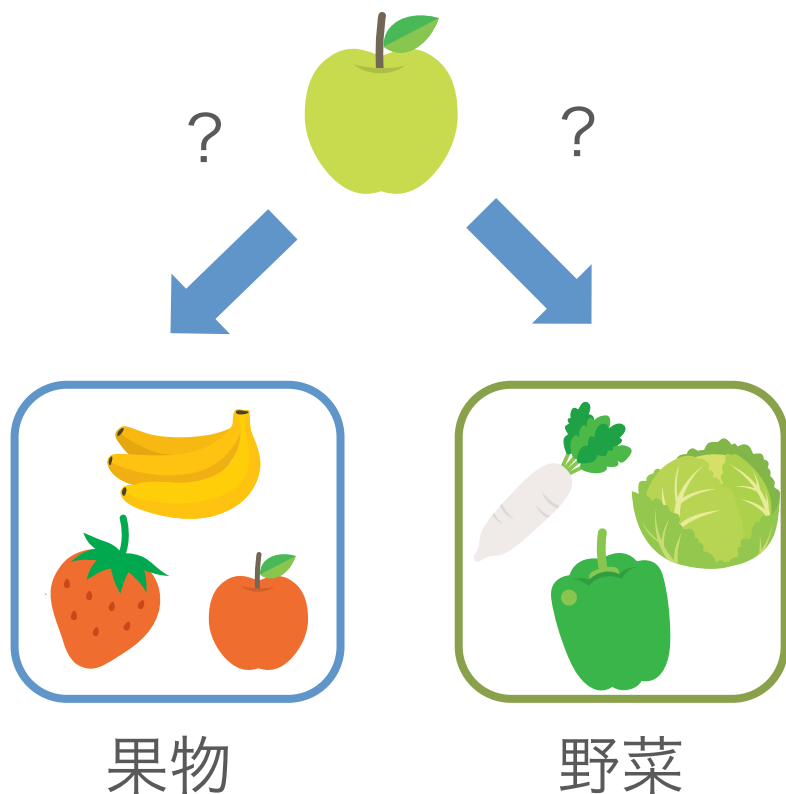
幅広い応用



機械学習の種類

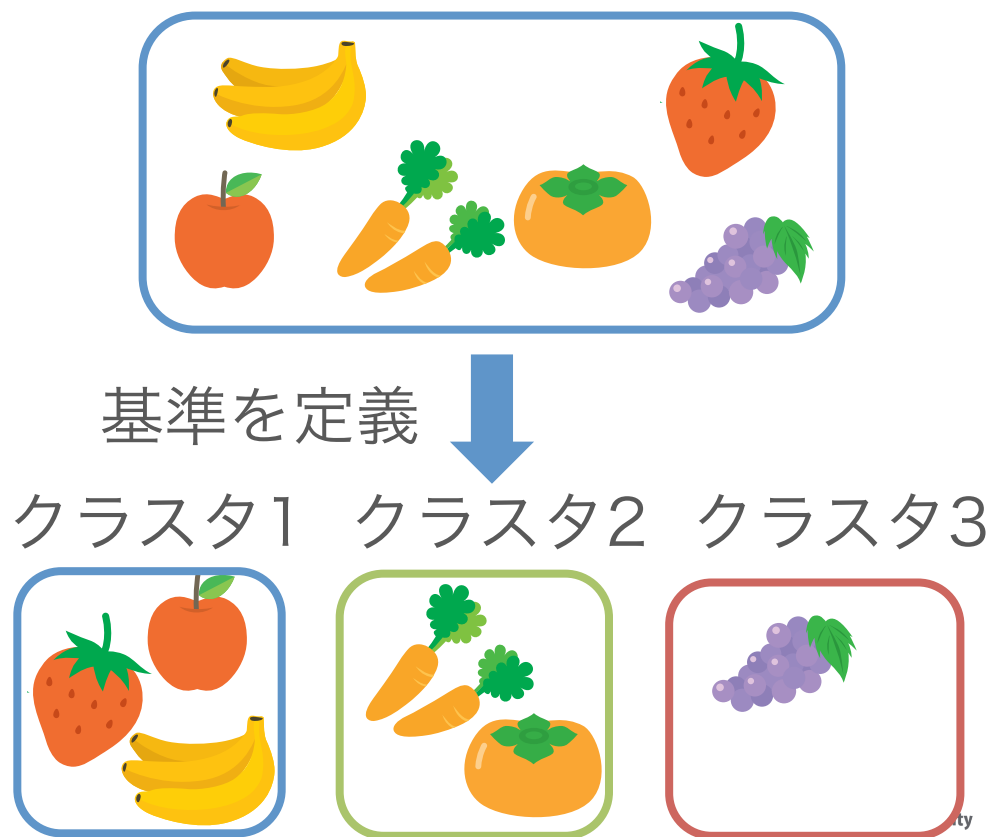
教師あり学習

データと正解のペアをもとに
分類モデルを生成



教師なし学習

特徴の似たものの同士の
グループ (クラスタ) を作る



目次

🌙 背景

🌙 従来の震源分類の妥当性の検証

🌙 サブクラスの有無の検証

🌙 まとめ



検証手順

波形から特徴量を抽出

Support Vector Machine (SVM)を適用

5交差検定の結果から従来の分類の妥当性を考察

SVMは教師あり学習のひとつ
手法の詳細は、次の発表者の菊池が説明



検証手順

- 従来の特徴量が、従来とは異なる特徴量を用いて、教師あり学習により再現可能か

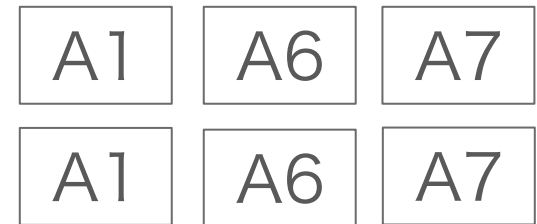
学習データ



分類器を生成



テストデータ



ラベルに誤りが含まれない



ラベルに誤りが含まれる可能性



検証手順

- 従来の震源分類が、従来とは異なる特徴量を用いて、教師あり学習により再現可能か

学習データ

分類器を生成



テストデータ

ラベルに誤りが含まれない

ラベルに誤りが含まれる可能性

A1

A6

A7

A1

A6

A7

特徴量

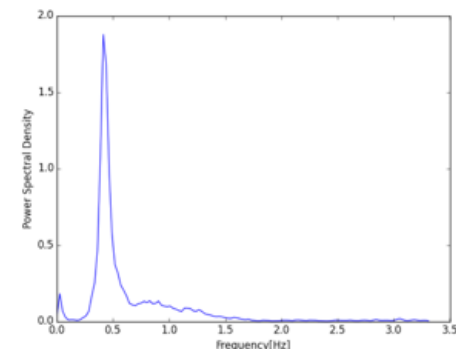
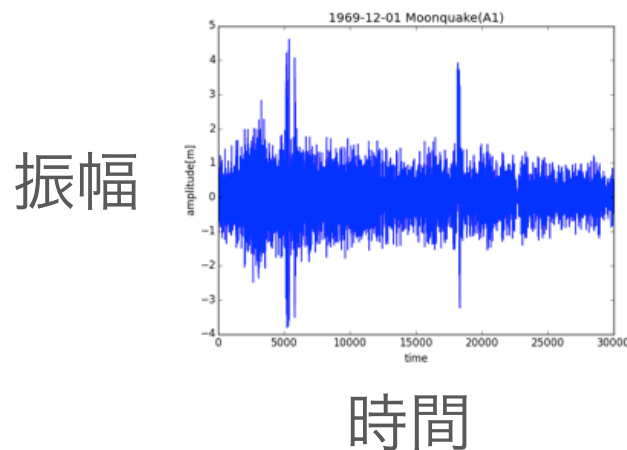
☾ 波形のPower Spectral Density (PSD) を使用

🌐 最終的な特徴量は、ベクトル長を1に正規化

☾ Power Spectral Densityとは

🌐 単位周波数あたりのパワー値

🌐 不規則信号の評価によく用いられる



データセット

- アポロ12号のデータを利用
- 人手でSN比の高い波形を選定

震源	A1	A6	A7	A8	A9	A10	合計
月震数	173	32	29	39	62	35	370

前処理

1. オフセット処理
2. トレンド引き
3. バンドパスフィルタ
(0.3~1.5Hz)
4. デスパイキング処理

各種パラメータ

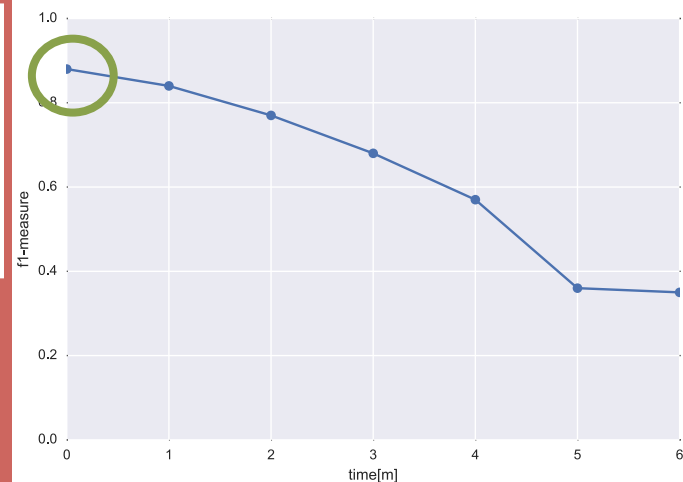
使用データ	LPZから約15分
セグメント長	512点
窓関数	Hanning
クラスタ結合法	Ward法



実験結果

P波到達時刻から、
切り出し開始時刻までの時間の差と
SVMの精度の推移

月震開始直後の波形に
震源ごとの特徴が現れやすい



	A1	A6	A7	A8	A9	A10
F値	0.95	0.89	0.93	0.85	0.95	0.72

☾ PSDは、震源分類を行う上で有効な特徴量

☾ 従来と異なる特徴量から見ても、
従来の震源分類に則した分類が可能

🌐 分類に失敗する理由：

○機械学習の限界 △ラベルに誤りが含まれる可能性

目次

🌙 背景

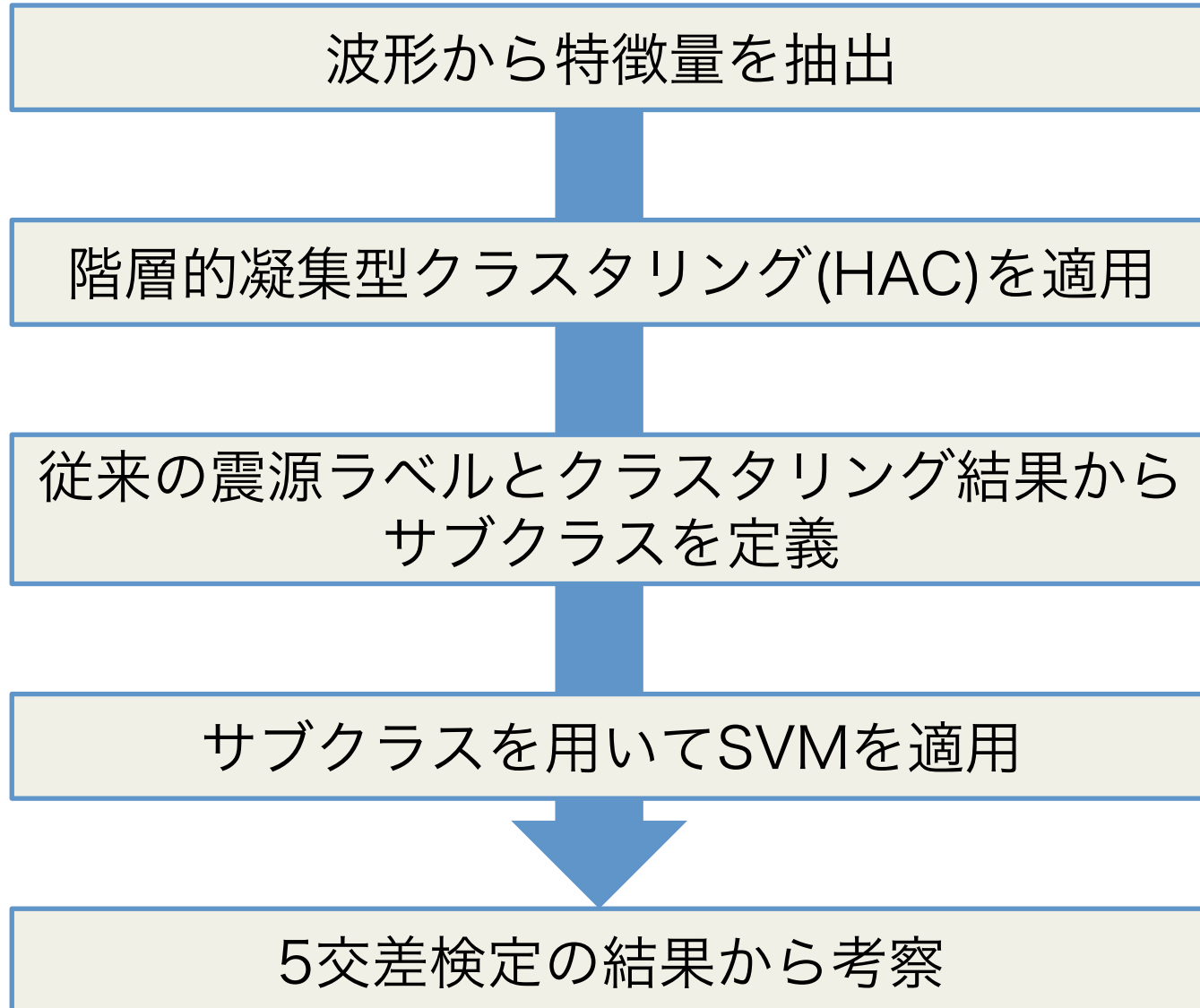
🌙 従来の震源分類の妥当性の検証

🌙 サブクラスの有無の検証

🌙 まとめ

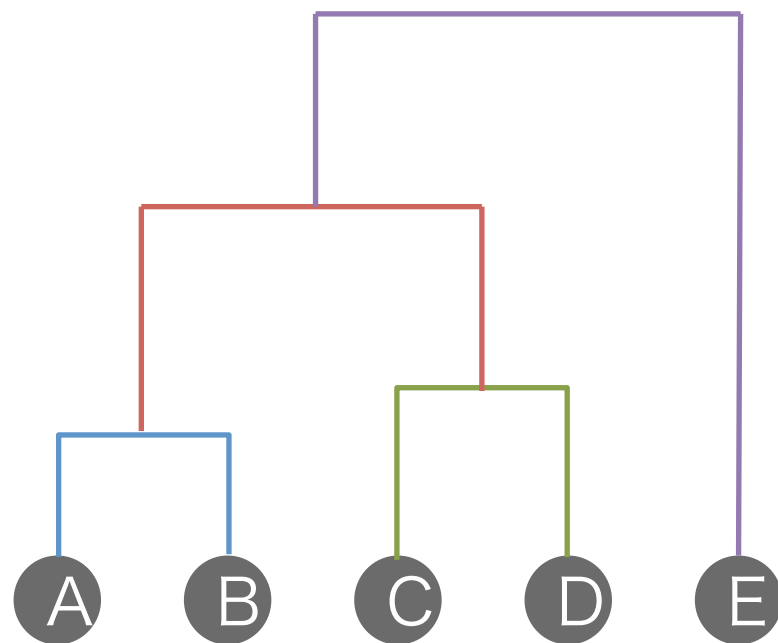
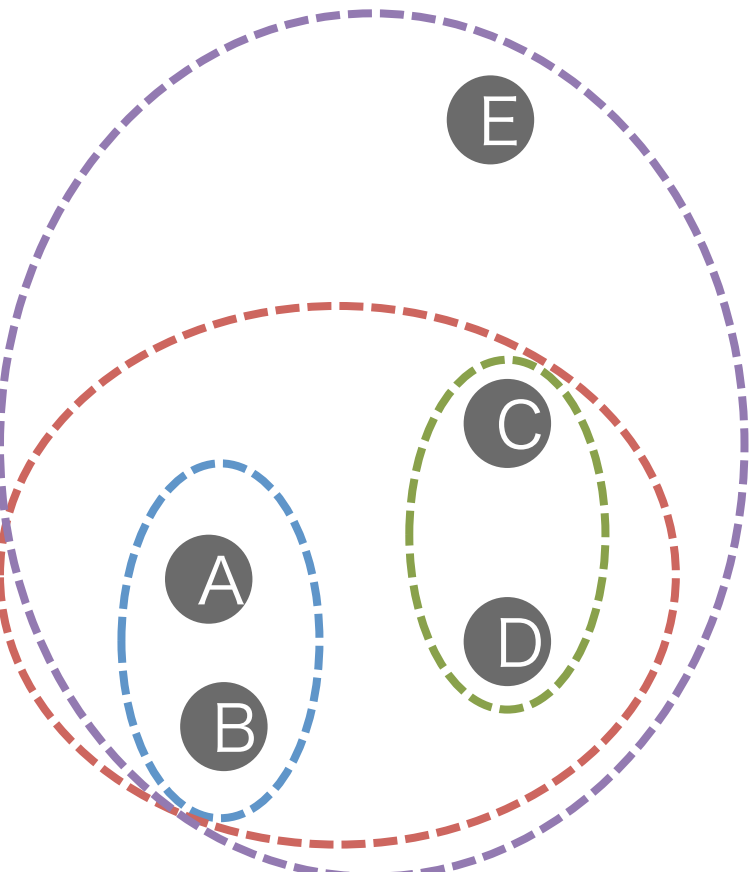


手法

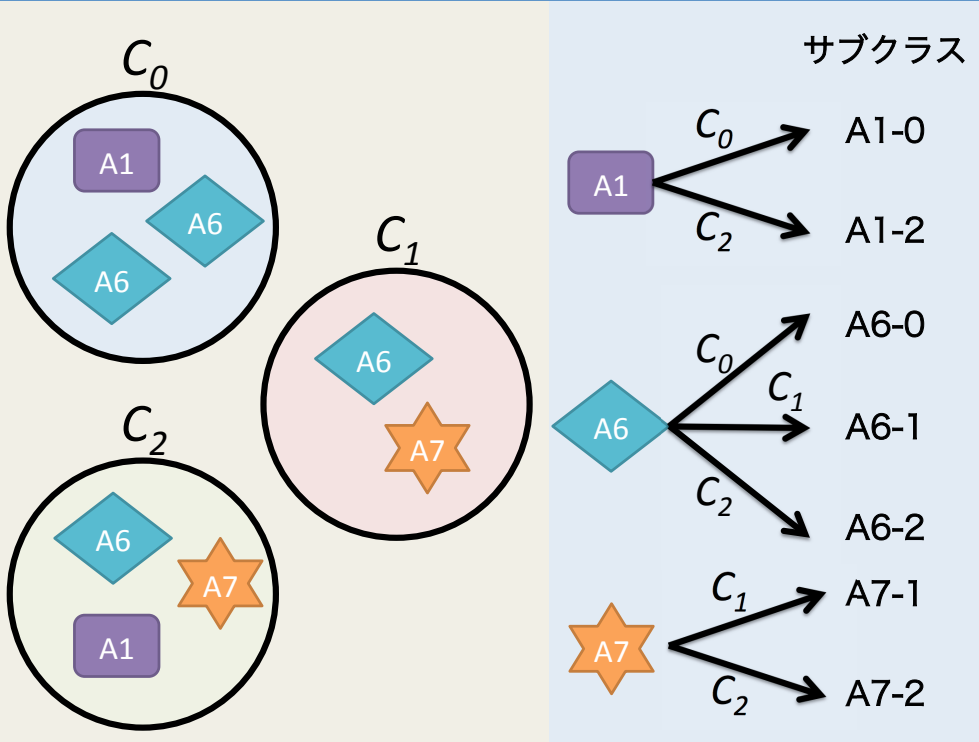


階層的凝集型クラスタリング (HAC)¹⁸

- すべてのデータが単一のクラスタとしてクラスタリング開始
- 距離関数に基づき、最も距離の近い2つのクラスタ併合
- すべてのデータが一つのクラスタになるまで逐次的に併合



手法



1. 月震をクラスタリング
2. クラスタリング結果と従来の震源ラベルを組み合わせ、サブクラスを定義
3. サブクラスを用いたSVMの交差検定でサブクラスとして成立するか検証

分類性能が高い：サブクラス固有の特徴がある
分類性能が低い：サブクラスとしては成立しないクラス

実験結果 (クラス数：10)

サブクラス	F値
A1-0	1.00
A1-6	0.60
A1-7	0.92
A1-9	0.63
A6-2	1.00
A6-6	0.78
A7-2	1.00
A7-8	0.96
A7-9	0.50

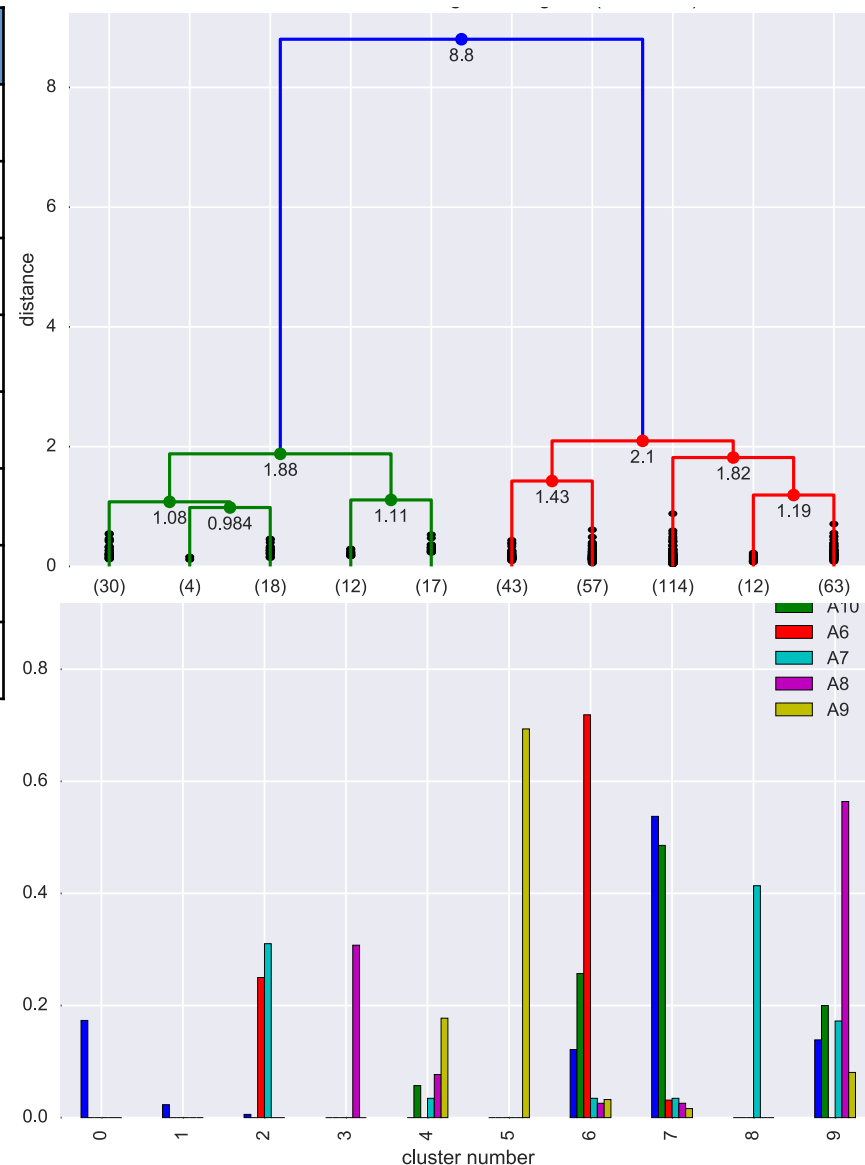
サブクラス	F値
A8-3	1.00
A8-9	0.76
A9-4	1.00
A9-5	0.92
A9-9	0.44
A10-6	0.67
A10-7	0.77
A10-9	0.22

左上：サブクラスごとのSVMのF値

右上：デンドログラム

右下：クラスタ内の震源別内訳

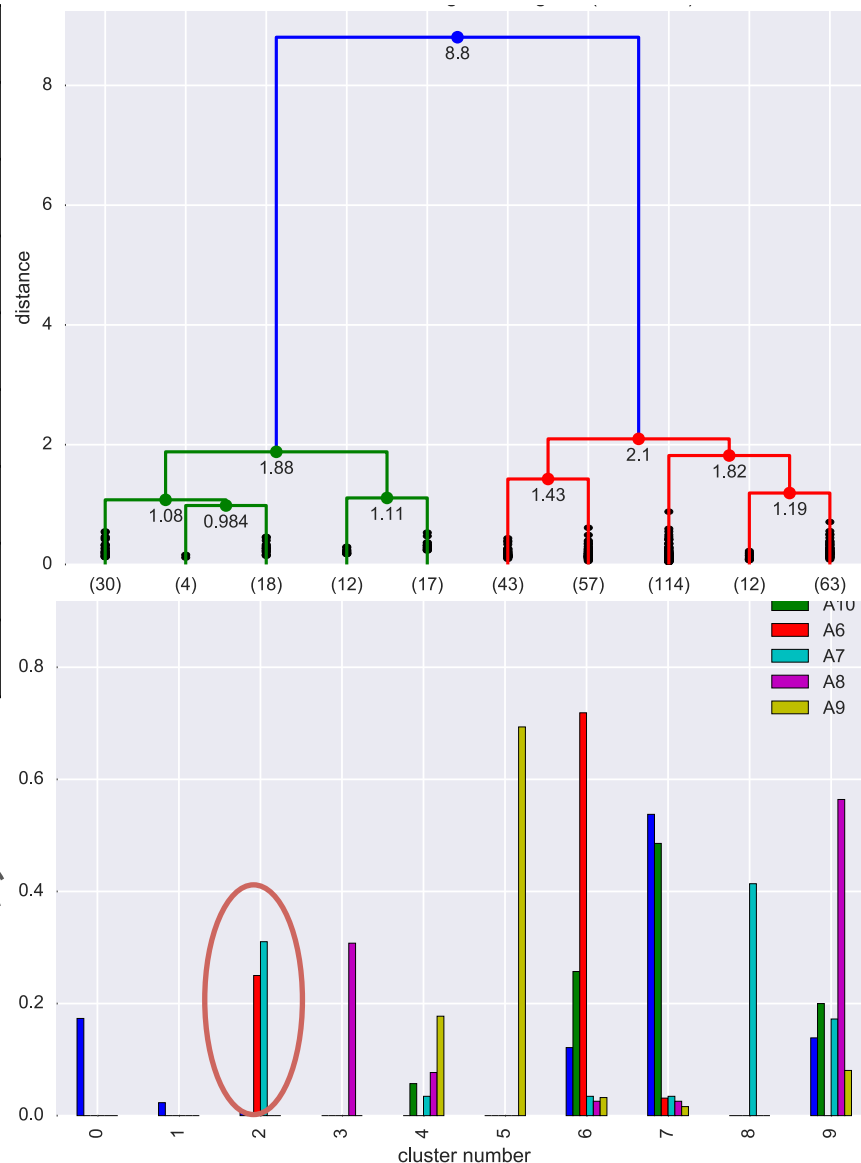
縦軸…ある震源のクラスタ内のデータ数
ある震源のデータ総数



クラスタに着目した考察

サブクラス	F値
A1-0	1.00
A1-6	0.60
A1-7	0.92
A1-9	0.63
A6-2	1.00
A6-6	0.78
A7-2	1.00
A7-8	0.96
A7-9	0.50

サブクラス	F値
A8-3	1.00
A8-9	0.76
A9-4	1.00
A9-5	0.92
A9-9	0.44
A10-6	0.67
A10-7	0.77
A10-9	0.22



同一クラスタに複数の震源があるが
分類性能が高い

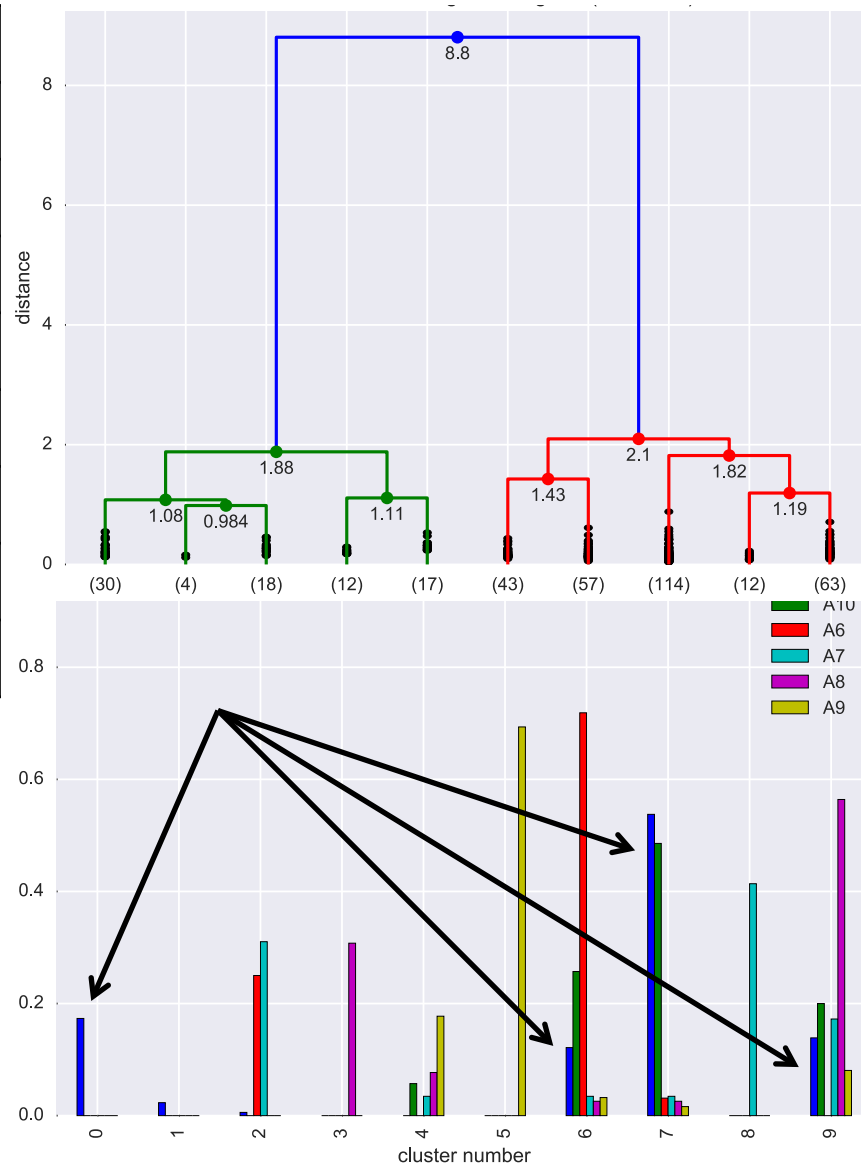
サブクラスとして成立

震源に着目した考察

サブクラス	F値
A1-0	1.00
A1-6	0.60
A1-7	0.92
A1-9	0.63
A6-2	1.00
A6-6	0.78
A7-2	1.00
A7-8	0.96
A7-9	0.50

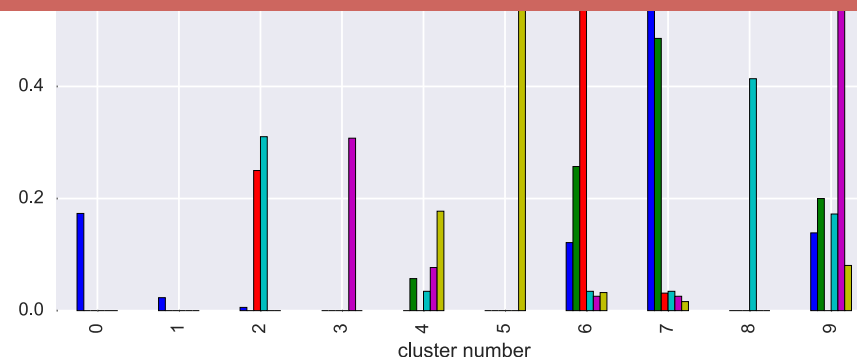
サブクラス	F値
A8-3	1.00
A8-9	0.76
A9-4	1.00
A9-5	0.92
A9-9	0.44
A10-6	0.67
A10-7	0.77
A10-9	0.22

- A1-0とA1-7は分類性能が高い
- A1-6とA1-9は分類性能が低い
→ サブクラスとしては不成立



- 同じ震源でも、複数の特徴が見られる
→ 分類を細分化可能である
- 月震の性質に近い震源が存在する

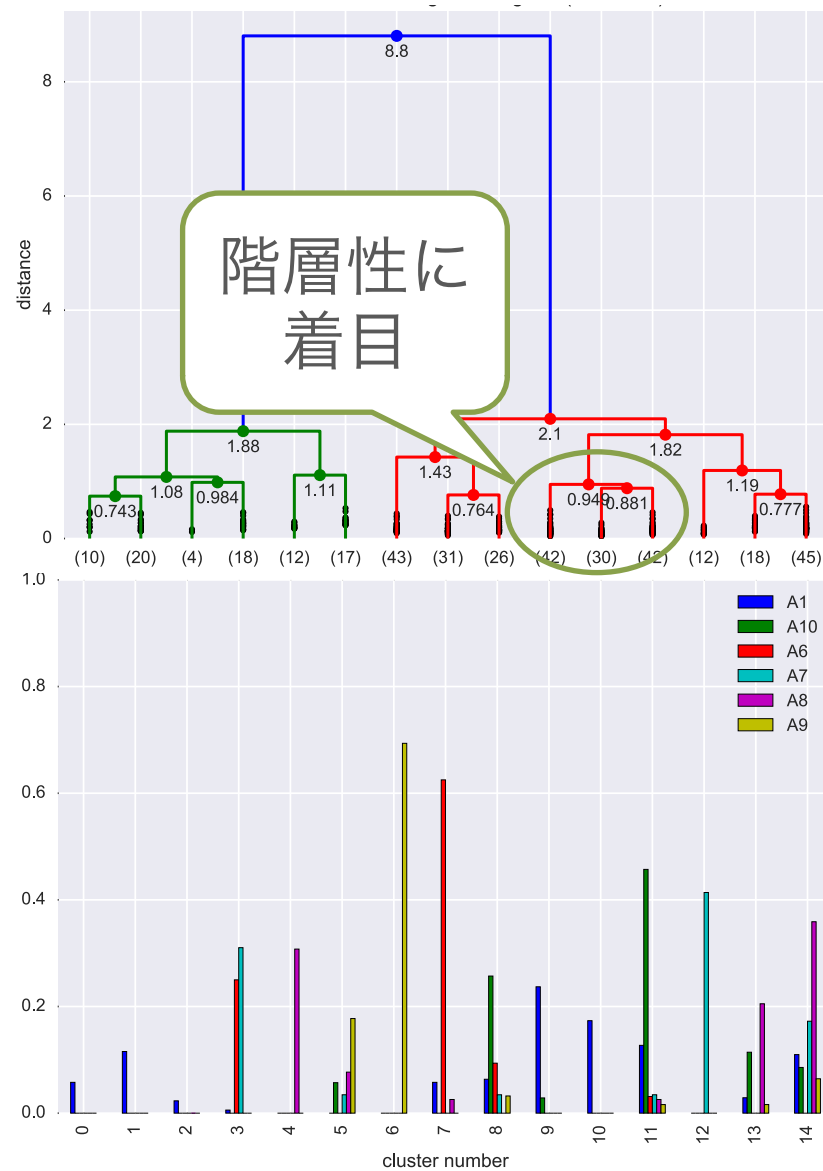
縦軸… $\frac{\text{ある震源のクラスタ内のデータ数}}{\text{ある震源のデータ総数}}$



実験結果 (クラス数 : 15)

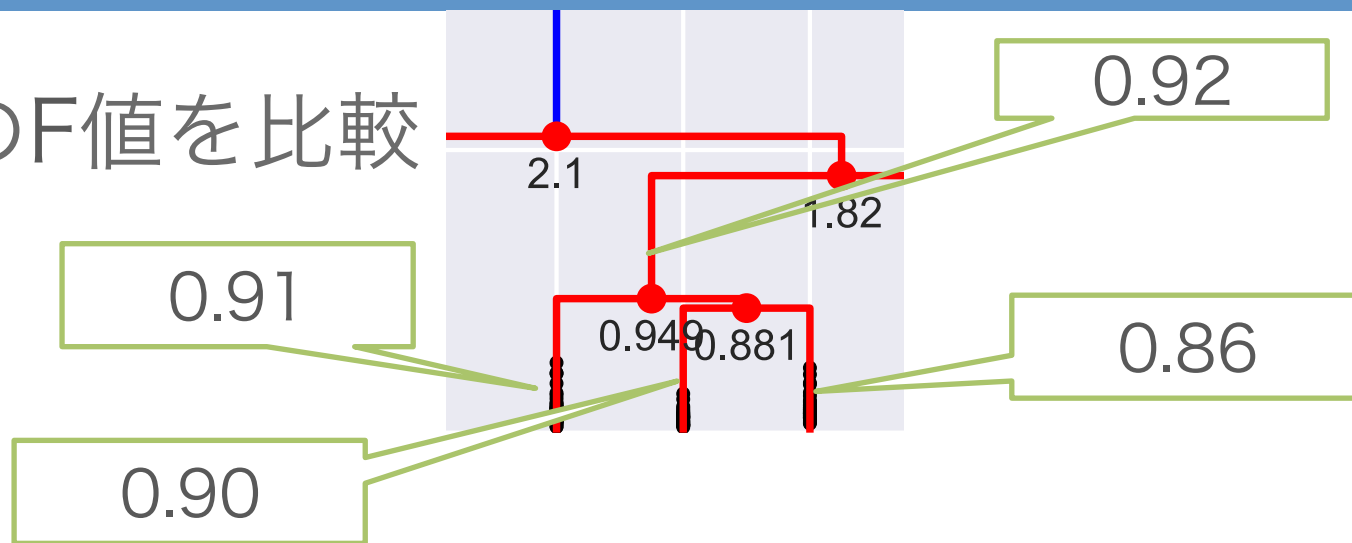
サブクラス	F値
A1-0	1.00
A1-1	1.00
A1-7	0.52
A1-8	0.78
A1-9	0.91
A1-10	0.90
A1-11	0.86
A1-13	0.60
A1-14	0.73
A6-3	1.00
A6-7	0.89

サブクラス	F値
A7-3	1.00
A7-12	0.96
A7-14	0.50
A8-4	1.00
A8-13	0.80
A8-14	0.89
A9-5	1.00
A9-6	0.94
A10-8	0.80
A10-11	0.83



階層性に着目した考察

A1のF値を比較



分割してもF値が高い
→ さらに細かく特徴ごとに分けられている

クラスタをさらに細分化することが可能
※どこまで分割するべきか、
閾値の決定などは今後の課題である

まとめ

- ☾ PSDは月震分類の特徴量として適切
- ☾ 従来と異なる特徴量から見ても、
従来の震源分類に則した分類が可能 → 妥当性がある
 - 🌐 ただし、従来の分類にラベルの誤りがある可能性
- ☾ 震源分類は細分化が可能
- ☾ 今後の課題
 - 🌐 デンドログラムの半自動的な分割法の考案
 - 🌐 他特徴量による比較

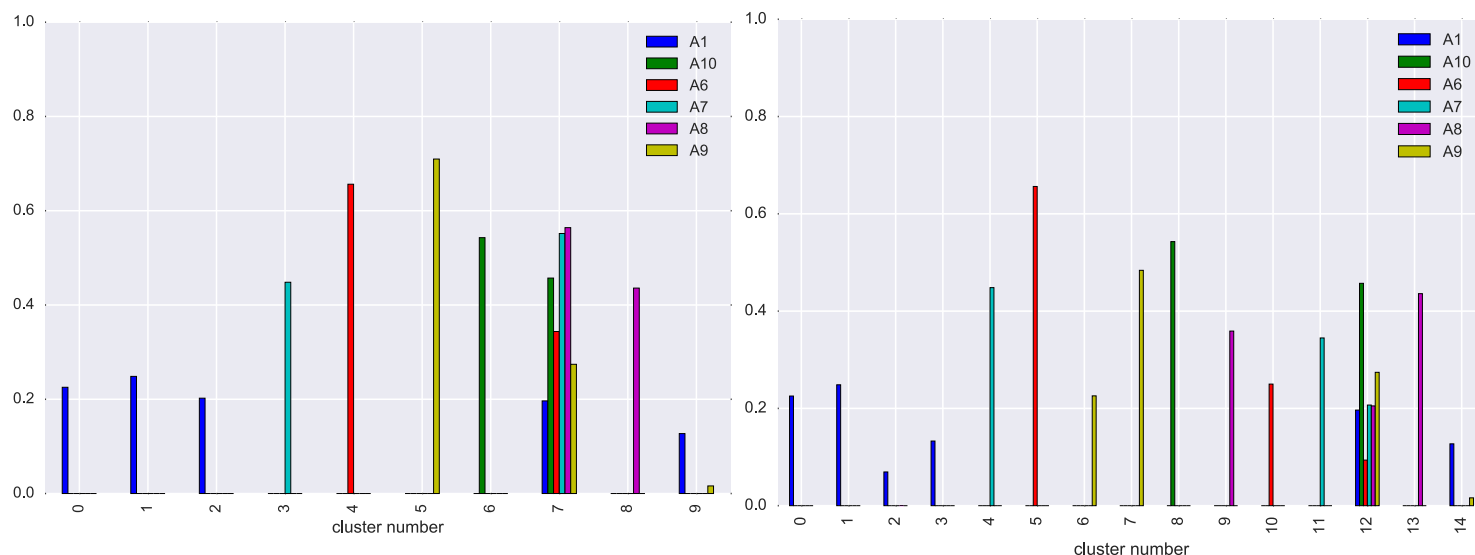


今後の課題

☾ 他の特徴量を用いた比較

🌐 今はPSDの性質に依存している

他の特徴量を用いて、震源の汎用的な性質を導出



図は波形のヒルベルト変換を用いたクラスタリング





交差検定とは

訓練データをk分割して精度を検証

Train	Train	Train	Train	Test
-------	-------	-------	-------	------

Train	Train	Train	Test	Train
-------	-------	-------	------	-------

Train	Train	Test	Train	Train
-------	-------	------	-------	-------

Train	Test	Train	Train	Train
-------	------	-------	-------	-------

Test	Train	Train	Train	Train
------	-------	-------	-------	-------

学習データに適応しすぎて、
汎化性能が失われることを
防ぐ目的

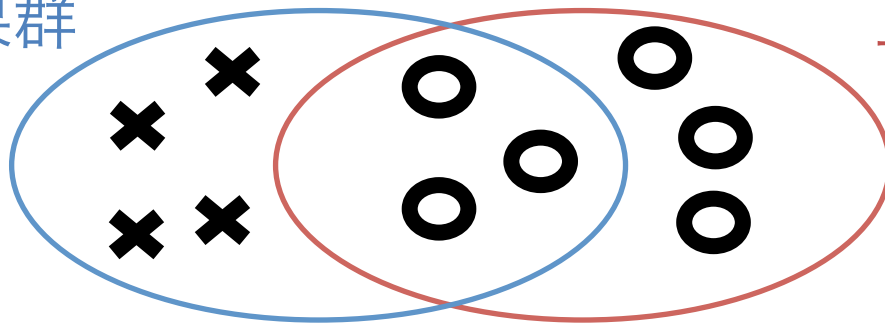
分類器の性能を評価する
一般的な手法



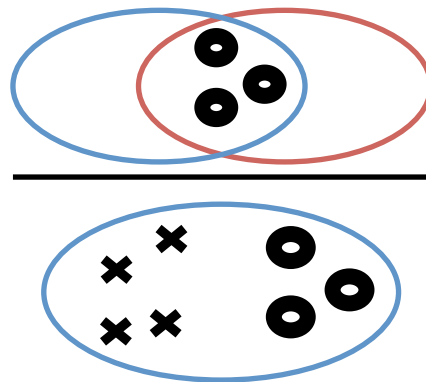
適合率, 再現率, F値とは

結果群

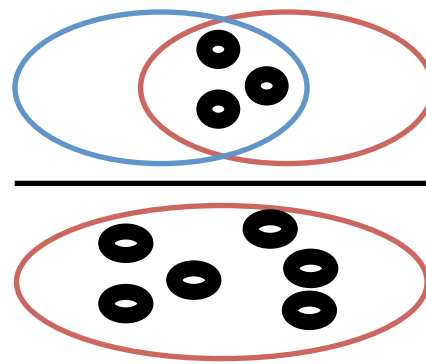
正解群



適合率 (Precision) =



再現率 (Recall) =



F値

$$\frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

教師あり学習を用いた検証

分類性能から考察できること

ラベルに誤りが含まれない

分類性能が高い



1. 特徴がうまく捉えられている
2. 人手で付与されたラベルが正しい

ラベルに誤りが含まれる可能性

分類性能が低い



1. 特徴がうまく捉えられていない
2. 人手で付与されたラベルに間違いが含まれる可能性がある

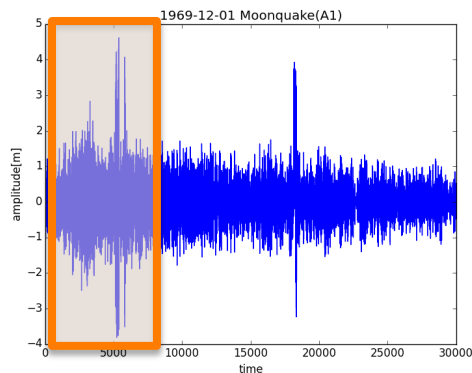


Ward法

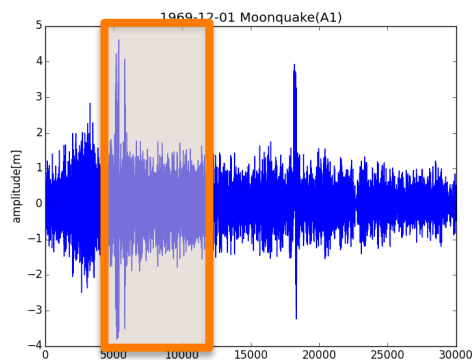
- ☾ 階層的クラスタリング手法の中では、安定した性質を持っていると言われる
- ☾ 階層法の中で最も精度が高い
- ☾ クラスタ結合の際に、重心からの距離の二乗和が最小となる2つのクラスタを併合



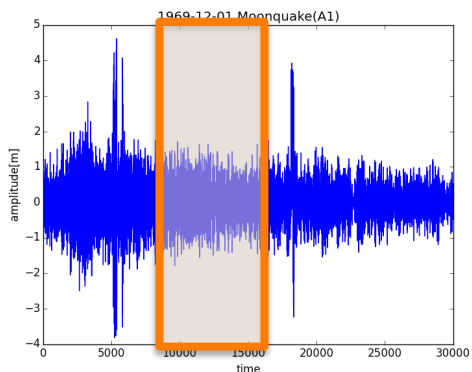
P波到達からの推移とは



☾ 特徴量を計算する場所を
シフトさせていく

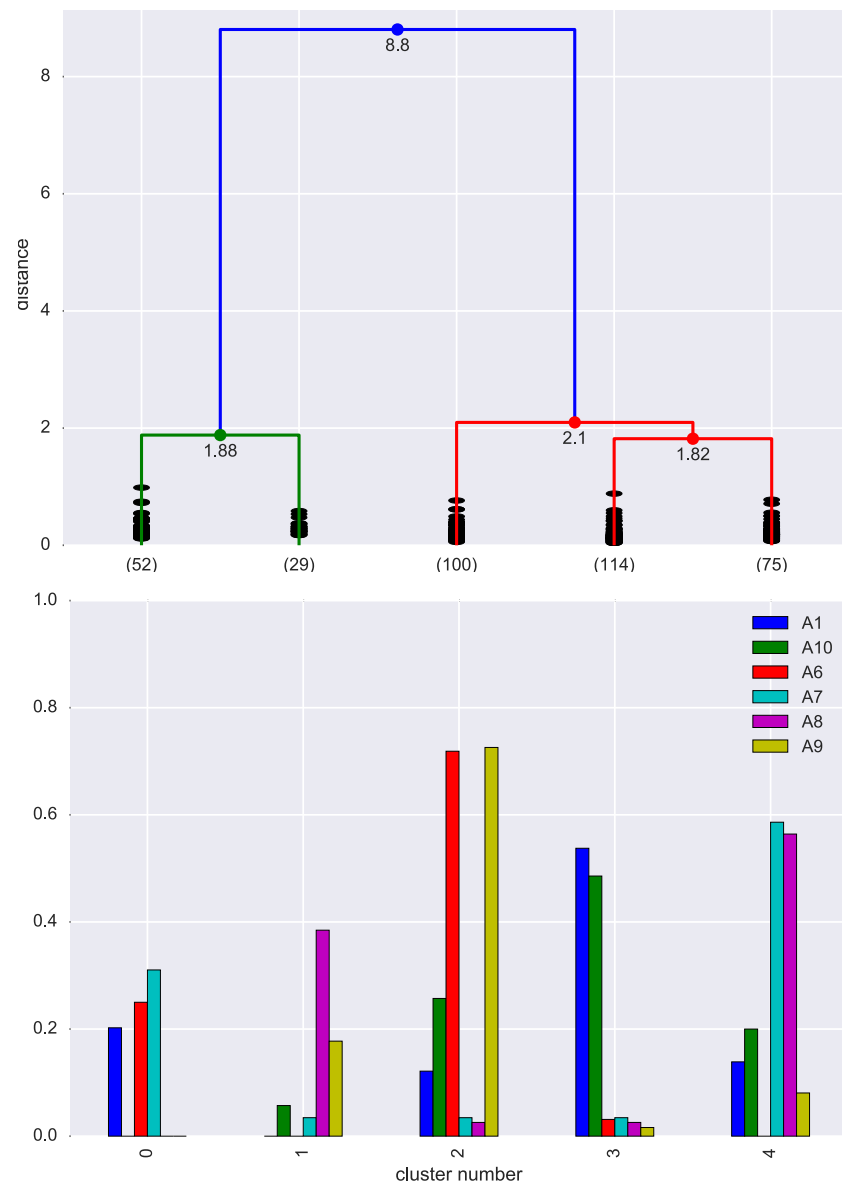


☾ P波到達時刻を0として、
スタート位置をずらしていく



実験結果 (クラス数 : 5)

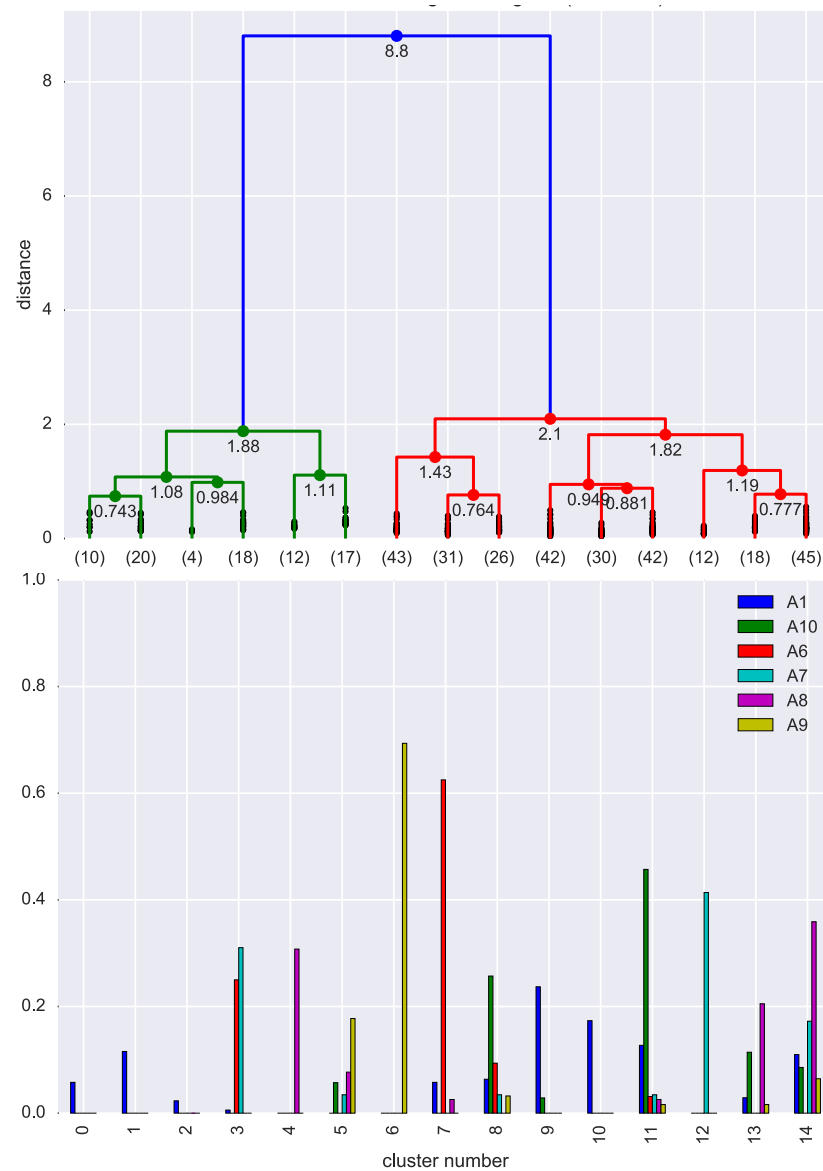
サブクラス	F値	サブクラス	F値
A1-0	0.97	A8-1	0.97
A1-2	0.59	A8-4	0.77
A1-3	0.95	A9-1	0.96
A1-4	0.58	A9-2	0.91
A6-0	0.88	A9-4	0.44
A6-2	0.86	A10-2	0.67
A7-0	1.00	A10-3	0.83
A7-4	0.91	A10-4	0.12



実験結果 (クラス数 : 15)

サブクラス	F値
A1-0	1.00
A1-1	1.00
A1-7	0.52
A1-8	0.78
A1-9	0.91
A1-10	0.90
A1-11	0.86
A1-13	0.60
A1-14	0.73
A6-3	1.00
A6-7	0.89

サブクラス	F値
A7-3	1.00
A7-12	0.96
A7-14	0.50
A8-4	1.00
A8-13	0.80
A8-14	0.89
A9-5	1.00
A9-6	0.94
A10-8	0.80
A10-11	0.83



これまでの月震研究

NASAの アポロミッション

月に地震計を設置
約7年半データ取得

月の地震の存在を確認

月震解析で得られる知見
発生要因, 内部構造,
震源 etc.

これまでの月震分類の内訳

深発月震	7245
浅発月震	28
熱月震	825(AP14) (313days)
隕石衝突	1744
人口月震	

[Nakamura et al., 1981,2003]
[Duennebier et al., 1974]

深発月震は震源でもさらに細かく分類



これまでの月震研究

🌙 NASAのアポロミッション

- 🌐 月に地震計を設置し,
月の地震の存在を確認
- 🌐 約7年半データ取得

🌙 月震解析で得られた知見

- 🌐 内部構造, 発生要因, 震源 etc.
- 🌐 月震の分類例

深発月震
熱月震

浅発月震
隕石衝突

etc.

半数以上が深発月震

月震の特徴

小さなS/N比

- ノイズが多く,
分類が困難

小さな振幅

- 深発月震の
エネルギーは
M0.5~1.3程度

揺れの継続時間

- 1時間以上揺れが
続くものも

