

修士論文

MFCC を用いたインフラサウンド波形の特徴量抽出および

機械学習による地球物理学的イベントの分類

Classification of geophysical events by machine learning and

feature extraction of infrasound waveform using MFCC

報告者

学籍番号: 1225065

氏名: 岡本 大輝

指導教員

山本 真行 教授

令和2年2月10日

目次

1	序論	3
1.1	インフラサウンドとは	3
1.2	背景	3
1.3	目的	5
2	火山噴火波形の MFCC による特徴量抽出と SVM による分類	6
2.1	インフラサウンドの観測手法	6
2.2	開発環境と教師あり学習について	8
2.2.1	開発環境	8
2.2.2	教師あり学習と性能評価値	8
2.2.3	線形識別関数と SVM	10
2.3	火山噴火波形のデータセット	11
2.3.1	桜島の火山噴火時刻の取得方法	11
2.3.2	訓練データにおけるデータ長の決定方法	11
2.3.3	低周波数ノイズ除去および火山噴火波形の最小周波数を求めるためのアルゴリズム	13
2.3.4	相互相関関数による他センサーが取得した火山噴火波形の選定	17
2.3.5	火山噴火波形以外のデータセット収集条件	22
2.4	ケプストラム分析と MFCC による特徴量抽出	23
2.4.1	ケプストラム分析	23
2.4.2	MFCC による音声の特徴量抽出	24
2.4.3	MFCC による動的に変化する場合におけるデルタケプストラムによる特徴量抽出	26
2.4.4	特徴ベクトルの構成	26
2.5	可聴音データセットとの対照実験	26
2.5.1	Speech Commands Dataset について	26
2.5.2	火山噴火波形データセットに合わせたデータセットの構成方法	27
2.5.3	Speech Commands Dataset と火山噴火波形のパラメータの比較	28
3	実験結果	29
3.1	気象庁から取得される火山噴火情報数	29
3.2	火山噴火波形の継続時間および訓練データのデータ長の決定	29
3.3	相互相関関数を用いて選定された各センサーのデータセット数	30
3.4	Speech Commands Dataset における特徴量抽出および分類	33

3.4.1	MFCC またはデルタケプストラムを特徴量とした性能評価値	33
3.4.2	データセット数を説明変数とした性能評価値の変化	35
3.5	火山噴火波形の特徴量抽出および分類	38
3.5.1	相関値の閾値倍率を説明変数とした性能評価値の変化	38
3.5.2	MFCC またはデルタケプストラムを特徴量としたときの最大性能評価値 .	41
4	考察・評価	42
4.1	データセット数の最適値	42
4.2	閾値変化によるデータセット破棄数の理論化	44
4.3	可聴音と火山噴火波形との比較	48
5	結論	55
	謝辞	56
	参考文献	57
	付録	60

1 序論

1.1 インフラサウンドとは

インフラサウンドとは可聴周波数下限の 20 Hz 以下の圧力波のことである (図 1.1)。インフラサウンドは可聴音と異なる特性として低周波数音波であるために長距離伝播できるということが挙げられる。これは音波は大気分子の粘性に支配されており、大気粘性には周波数依存性があるためである [1]。インフラサウンドも音波であるため空気中の様々な要因で発生することがあるが、その波長の大きさから大規模自然災害をもたらす地球物理学的イベント、すなわち火山噴火、津波、地震、台風などが発生源となる場合がある。



図 1.1 可聴音に対する音波の分類

1.2 背景

近年では、CTBT(包括的核実験禁止条約)に基づく核実験の監視装置の一つとして微気圧振動、すなわちインフラサウンドを用いた監視が行われている。国内における CTBT の微気圧振動観測所は日本気象協会が千葉県いすみ市に観測点 (IS30) を運用しているのみである。同観測点では 2011 年 3 月 11 日に発生した東日本大震災の大津波が励起したときに発生したインフラサウンドの到達が確認された [2]。その圧力波の到達時刻は津波到達時刻より早く観測されたことから、音波伝播速度は津波の平均進行速度より速く津波波源域の情報を伝達できることが示された。そのため従来の津波予測とは異なり、陸上設置のインフラサウンドセンサー群により津波をリモートセンシングするアプローチで津波到達を把握できると期待されている。

当初のインフラサウンドセンサーは本体価格が高く、さらに 2000 年代には輸入品しかなかったために、密なインフラサウンドセンサーネットワークを構築するためには安価なセンサーの開発が求められた。本研究室では、西山 (2007) による広い周波数特性に圧力感度を持つ piezo 素子を用いた低価格なセンサー本体の基礎開発 [3]、山田 (2009) により PSD(Position Sensitive Device) と半導体レーザーを用い、振動膜面を非接触センシングする方式の基礎開発が行われ、周波数下限を 0.001 Hz まで広げることができ、センサー本体の価格も抑えことに成功した [4]。これらの成果を生かし、高知工科大学と株式会社サヤの

共同開発により、2015年にはインフラサウンドだけでなく、地震に対する3軸加速度計、騒音計、気圧計、温度計を搭載した複合型インフラサウンド津波センサー（以下、インフラサウンドセンサー）が発表された。2016年には高知県黒潮町に同センサーを5台を設置、2017年には高知県内沿岸域の計15か所に拡充し、津波起因インフラサウンドの到達を面的に検知するネットワークが構築された[5]。現在でもインフラサウンドセンサーの増設が行われており、これまでに合計16か所に設置されている（図1.2）。インフラサウンドセンサーを活用した防災・減災に関する重要な課題の一つとして、大規模自然災害の発生時にアラートをリアルタイムで発令するシステムの構築があり、反町（2016）によってN形波形イベントを検出するアルゴリズムの基礎開発が行われた[6]。N形波形とは突発的な現象の際にインパルス的に現れる波形のことであるが、火山噴火、雷といった大規模地球物理学的イベントだけではなく、花火や屋内設置の場合はドアの開閉などの現象でも発生する。そのため、時系列データをスキャンさせてN形波形を検知することは可能であるが、波形から音源の種類を特定することには至っていない。

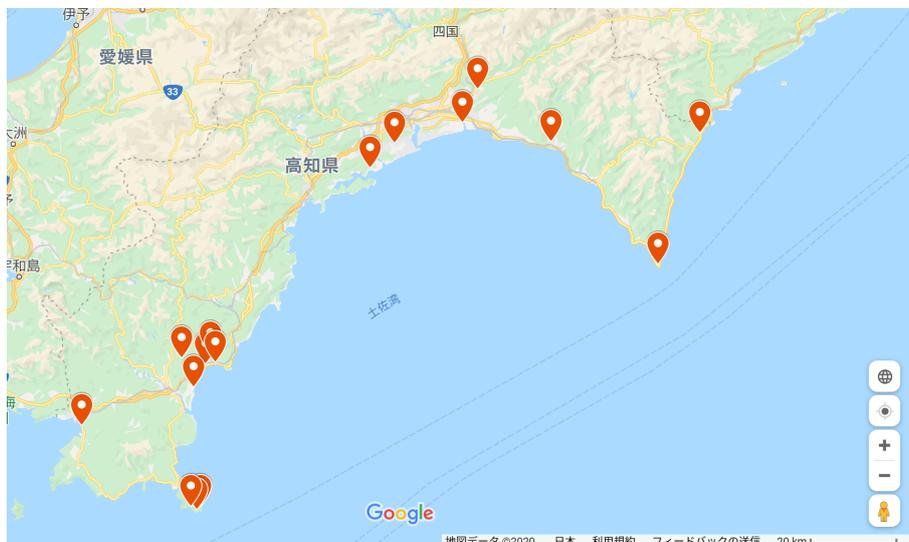


図 1.2 高知県におけるインフラサウンドセンサー群の配置箇所。各センサーの設置箇所が赤ピンで示されている。

本論文執筆時点において、観測開始から現在まで高知県沿岸部での津波は発生していないため津波の波形データは存在しない。しかしインフラサウンドが発生する大規模自然災害の一つとして火山噴火が挙げられる。特に鹿児島県の鹿児島湾内に位置する桜島の火山噴火による衝撃波はインフラサウンドとして伝播し、高知県に設置されたインフラサウンドセンサー群が特徴的な波形データを取得していることが確認できる。ここで桜島における火山噴火の発生頻度について、気象庁の発表では2014年1月7日から2019年11月3日の期間に東郡元空振計で1515回の観測がされている[10]。これは平均すると2日に1回ほどの噴火頻度であると計算され、桜島は頻繁に噴火していることが確認される。さらに面的配置されたインフラサウンドセンサー群が定点観測を開始した2016年から火山噴火波形の到達回数

を考慮すると、十分な波形データが取得できていると考えられる。

ところで、近年計算機の発達により機械学習に大きな注目が集まっている。機械学習の手法の一つとして教師あり学習をさせることにより未知のデータを予測、分類する手法がある。その応用として、訓練データとして波形データを与えて学習させ、膨大なデータの中から特定の信号を発見すると手法が用いられることがある [11]。この方法を桜島の火山噴火によるインフラサウンド波形に適用することができれば、大規模自然災害が発生したときにアラートを自動発出する手段になると考えられる。また訓練データを集められたとしても機械学習のアルゴリズムを用いて音声信号の種類を分類するためには特徴ベクトルが必要である。そこで可聴音に対する特徴量抽出に注目すると、音声工学では非統計的な手法として声道成分と基本周波数成分を分離するケプストラム分析があり、低次元数に圧縮したメル周波数ケプストラム係数 (以下、MFCC) がある。さらに MFCC において非定常な周波数を持つ音声信号の特徴量を抽出できるデルタケプストラムがある [12]。この 2 点は音声工学において特徴量抽出の方法としてしばしば用いられる。近年では干ばつや切れ目によってストレスを受けた植物が発する超音波について、ストレスの種類で周波数特性が異なるため、MFCC によって特徴量抽出を行い機械学習アルゴリズムで分類する検証が行われた [13]。このように可聴音以外の周波数領域でも MFCC が用いられている。またインフラサウンドセンサー群がアレイ配置され、膨大な大規模自然災害の波形データが得られる中で、音声工学の特徴量抽出の基礎を適用することは、機械学習による分類を用いた防災・減災アラートシステムの構築を行うという意味で重要なアプローチだと考える。

1.3 目的

本研究では、大規模自然災害をもたらす地球物理学イベントにより発生したインフラサウンド波形の到達をトリガーとするリアルタイム防災・減災アラートの基礎開発を目指す。その具体的な手法として、膨大なインフラサウンド波形で学習させた分類器を用いて時系列データを任意の窓幅でスキャンさせることで到達した波形の種類を判断させる手法を提案する。またデータセットの収集の際、インフラサウンドセンサー近傍における予測困難な気象要因のために火山噴火波形が到達しなかった場合に対処するためのデータ選定のアルゴリズムの開発を目的とする。さらに、可聴音領域において MFCC およびデルタケプストラムの特徴量抽出し分類する手法がインフラサウンドデータを用いた大規模自然災害の分類にも適用できるかを検証する。

2 火山噴火波形の MFCC による特徴量抽出と SVM による分類

2.1 インフラサウンドの観測手法

インフラサウンドに限らず音波のデータは気圧変動を観測することによって得ることができる。しかし長周期のインフラサウンドの確実な検出には通常の気圧計に対し 1 mPa 程度までの微小気圧変動をとらえる必要があるため、専用のセンサー開発が求められる。高知工科大学と株式会社サヤが共同開発したインフラサウンドセンサーの ADXII INF01(図 2.1) および ADXIII INF01LE(図 2.2) では気圧変動の最小分解能がそれぞれ約 0.18 mPa、約 0.19 mPa となっている。また周波数範囲はどちらも 0.001 Hz ~ 6.25 Hz であるため、津波起因のインフラサウンドなどの長周期のイベントを捉えることができる [14][15]。



図 2.1 高知工科大学と株式会社 SAYA と共同開発した複合型インフラサウンド津波センサー ADXII INF01[16]



図 2.2 高知工科大学と株式会社 SAYA と共同開発した複合型インフラサウンド津波センサー ADXIII INF01LE

特に ADXIII INF01LE の仕様について説明する。ADXIII INF01LE はセンサ内部でインフラサウンド信号が AD 変換され、インフラサウンドを含む複数種類の計測データや GPS 時刻情報が統合され、予めインストールされた専用ソフトウェアによってリングバッファもしくはポーリング形式でイーサネットを経由しコンピュータに観測データが記録される。その後、4096 サンプルのデータが蓄積されると CSV 形式ファイルとして整形され、インターネットを介し FTP (File Transfer Protocol) で専用のデータサーバーに自動送信される。本研究では気圧変動となる時系列データを解析する。書き込まれた CSV 形式ファイルには GPS 時刻とサンプリングしたときのコンピュータ内部時刻のタイムスタンプが付加されている。今回の解析では時間の間隔が重要であり時刻精度については求めない。GPS 時刻を用いてもよいが、GPS 信号の受信状況に依存して時刻の重複したデータが存在するな

どの問題があった。そのため、UNIX 時刻で記録されたコンピューターの時刻を用いることにした。気圧変動についてはセンサー本体に搭載されているデジタルフィルタによるハイパスフィルタが施されたインフラサウンド AC データを用いた。ADXIII INF01LE でサンプリングされたデータについて、コンピュータの OS からのポーリングによる非同期サンプリングの結果、サンプリングレートを整数で定めることができない。しかし、リングバッファモードはこの限りではない。十分なデータが得られたときのサンプリングレートの平均は約 5 Hz である。ここでサンプリングレートが一定であると訓練データの固定データ長に定めることができる利点がある。サンプリングレートを一定にする手法 (リサンプリング) を図 2.3 に示す。

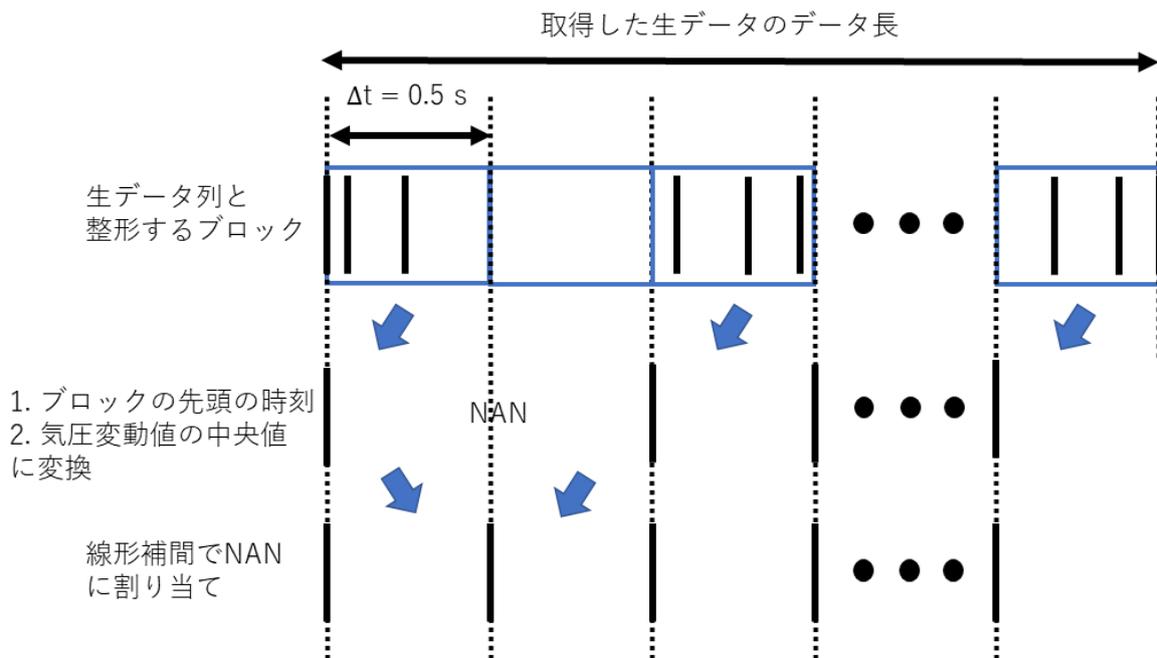


図 2.3 非同期サンプリングのデータのサンプリングレートを一定にする手法

始めにサンプリングレートを定める。今回の場合、サンプリングレートを 2 Hz とし、 $\Delta t = 0.5$ 秒ごとに始めと終わりの時刻情報を持つブロックを生成する。センサーから得られる時刻に対し、ブロックの時間内であればサンプリングデータをブロックの中に挿入する。その後、ブロックに入った気圧変動値それぞれの中央値を取り、ブロックの先頭の UNIX 時刻を選択し、時刻と気圧変動値の値を一意に定める。このようにすることでサンプリングレートを一定にすることができる。しかし、非同期サンプリングのためブロックに挿入されない時間もある。その場合には一度 NAN を割り当て、線形補間によって時刻と気圧変動の値を割り当てることでブロック内にデータがないことを回避する。

2.2 開発環境と教師あり学習について

2.2.1 開発環境

今回機械学習における分類のアルゴリズム、性能評価は scikit-learn と呼ばれる BSD ライセンスの Python 上で取得できる機械学習のライブラリを用いる [17]。scikit-learn は教師あり学習と教師なし学習における分類、予測のアルゴリズムだけでなく、データの前処理、モデルの選択と評価など機械学習全般のツールがサポートされている。使用した scikit-learn のバージョンは 0.19.1 である。開発言語は Python であり、Python のバージョンは 3.6.5 を使用した。

2.2.2 教師あり学習と性能評価値

教師あり学習は訓練データに基づいて機械学習アルゴリズムのモデルを構築し、未知のデータに対して予想する。訓練データと同じ性質を持つ未知のデータに対して、正確に予想できる場合、汎化性能があるという。モデルの性能を評価をするときは、汎化性能を計測するために訓練データとテストデータに分割し、テストデータで性能評価する [18]。構築したデータセットに対して訓練データとテストデータに分割して性能評価を計測することを一度しか行わない場合はその訓練データを過信することになる。そのため、本研究では統計的に汎化性能を評価する手法の k 分割交差検証法 (以下、交差検証法) を用いる [19]。交差検証法は、データを k 分割し、 k 回性能評価値を計測する方法である。 $k = 5$ のときの交差検証法を図 2.4 に示す。

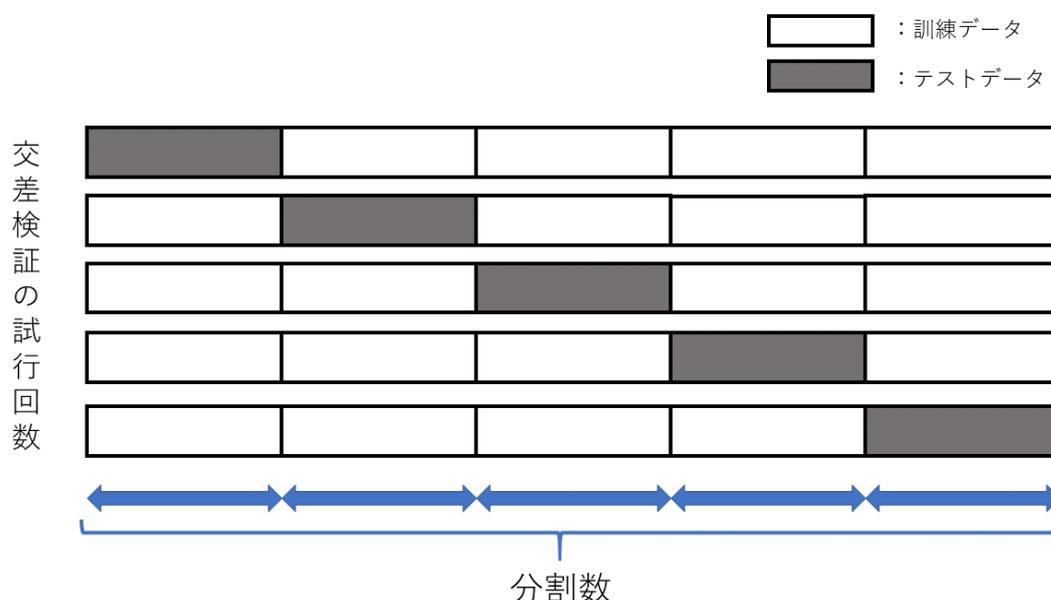


図 2.4 データセットを $k = 5$ 分割した場合の交差検証法

図 2.4 の場合では用意したデータセットを 5 分割するため、20% をテストデータ、80% を訓練データに分けることができ、テストデータの性能評価値を計測する。次に前回の訓練データと被らないように同じ割合で訓練データ、テストデータを割り当て、テストデータの性能評価値を計測する。この操作を 5 回繰り返し性能評価値を得る。ここでデータセット数 N から訓練データとテストデータに分ける方法は本来 N 通りである。そのうち $k = 5$ 通りを標本として抽出したと解釈すると、標本データから計算した標準偏差を s とした場合、標準誤差を用いて平均値のばらつきを考えることができる。本研究では交差検証法の 5 分割に対してデータセット数 N は十分大きい仮定し、標準誤差は $\frac{s}{\sqrt{k}}$ を用いる。本論文では標準誤差は 68% の信頼限界として表記する。

本研究の目的は取得されたセンサーの時系列データに対して任意の窓幅をリアルタイムでスキャンし、火山噴火波形かどうかを識別し、自動的にアラートを発令するシステムの構築である。そのため、分類タスクとして最も単純なモデルは火山噴火波形とノイズ波形を判定する 2 クラス分類である。ここで、2 クラス分類に本研究で用いる性能評価値について述べる。訓練データとテストデータをラベリングし、訓練データから学習した分類器のモデルを用いてテストデータでクラスを判定することを考える。分類器が判定したラベルは識別ラベルと呼ばれ、2 クラス分類の場合は陽性： p 、陰性： n が定義できる。また訓練データとして与えた値を真ラベル $\{p^*, n^*\}$ と表すことにする。今回の場合は火山噴火波形が p^* であり、それ以外の波形が n^* とする。分類器の判定は識別ラベル $\{p, n\}$ することができる。また真クラスと識別ラベルの組み合わせから得られる真陽性 (TP)、真偽性 (TN)、偽陽性 (FP)、偽陰性 (FN) を表 2.2.2 に示す。真クラスと識別ラベルの組み合わせから得られる行列のことを混合行列と呼ぶ [20]。それらから得られる性能評価値の計算式を式 (1) から式 (4) に示し、本研究でこれらの性能評価値を用いる理由を示す。ただし、真クラス p^* 、 n^* のデータ数をそれぞれ $P = TP + FN$ 、 $N = FP + TN$ と表す。

表 2.1 2 クラス問題の混合行列

		識別ラベル	
		p	n
真クラス	p^*	真陽性 (TP)	偽陰性 (FN)
	n^*	偽陽性 (FP)	真陰性 (TN)

$$\text{正解率 (accuracy)} = \frac{TP + TN}{P + N} \quad (1)$$

$$\text{再現率 (recall)} = \frac{TP}{P} \quad (2)$$

$$\text{適合率 (precision)} = \frac{TP}{TP + FP} \quad (3)$$

$$F\text{-値} = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} \quad (4)$$

- 正解率 (accuracy)

真ラベルに対して正しく分類されているかを見る性能評価値であり、最もよく使われる。しかし、分類器により TP または TN と判定されたテストデータが極端に多い場合は、 FP または FN が無視され、正解率が正しく出ない場合がある。この場合は、真ラベル p^* 、 n^* のデータ数 P 、 N を不均衡データにしないこと (付録) で解決できる。そのため、本研究では与えたラベルの比を 1:1 とした。

- 再現率 (recall)

真ラベルの中に真陽性が含まれる割合を表す。偽陰性の含有率の低さを見ているため、網羅性などと呼ばれている。今回のように火山噴火波形到達の自動アラートを出したい場合は、いかにして火山噴火波形の検知に失敗しないかと考えられ、開発者がユーザーにどれだけ信頼できるアラートシステムが提供できているかを示す度合となる。

- 適合率 (precision)

識別ラベルが陽性のデータに対して、真陽性が含まれる割合を表す。偽陽性の含有率の低さを見るため、正確性などと呼ばれている。火山噴火波形到達の自動アラートの場合、火山噴火波形が確認できないにも関わらず、アラートとして動作してしまった、すなわち誤報である考えられる。つまり誤報をユーザーに提供してしまう可能性の度合であると解釈できる。

- F-値

一般に適合率と再現率の間にはトレードオフの関係があるので、両者の調和平均をとることで関係性を評価する。その値を F-値という。

2.2.3 線形識別関数と SVM

本研究では 2 クラス分類を行うが、今回は線形識別関数が用いられたアルゴリズムを利用する。線形識別関数は識別境界を生成するための関数であり、入力データの次元を d とすると、 $d - 1$ 次元の超平面となる。一般に d 次元の入力ベクトルを $x = (x_1, \dots, x_d)^T$ 、係数ベクトルを $w = (w_1, \dots, w_d)^T$ 、バイアス項を w_0 とするとき、線形識別関数は $f(x) = w^T x + w_0$ と表される。ここで T は行列の転置である。線形識別関数における係数ベクトル w とバイアス項 w_0 を統計的に十分な量の訓練データを用いて学習させ、識別境界を決定する。識別境界を決定するアルゴリズムについて、本研究で用いるのはサポートベクトルマシン (以下、SVM) とした。SVM の概要を図 2.2 に示す。

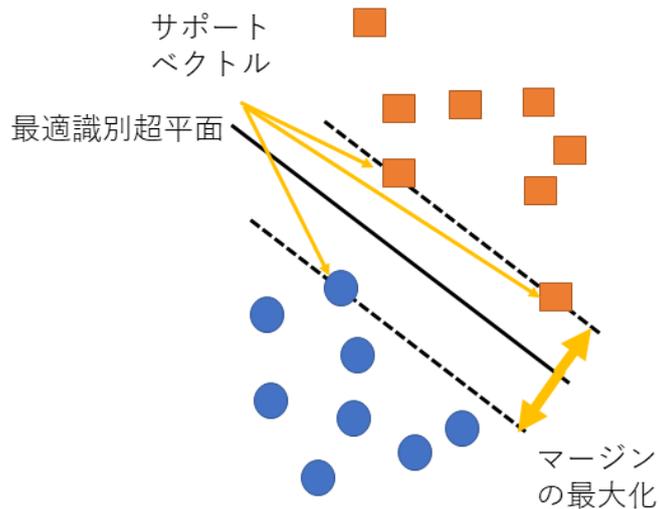


図 2.2 2次元平面における SVM の概要図

SVM は境界に最も近いサンプルとの距離 (マージン) を実現する 2 クラス線形識別関数の学習法である。ここで最適識別超平面を構成する要素のことをサポートベクトルといい、SVM はサポートベクトルを用いてマージンが最大になるように分類境界を選ぶアルゴリズムである [21]。SVM は線形分離可能でない学習データにおける対処方法があるが、本研究ではその手法は用いず、今回用意するデータセットは線形分離可能であると仮定する。

2.3 火山噴火波形のデータセット

2.3.1 桜島の火山噴火時刻の取得方法

気象庁が行っている火山噴火の観測手法の一つとして空振計による空振観測がある。空振観測は、天候不良などにより監視カメラで火山の情報監視できない場合でも噴火の規模を計測することができる [7]。特に気象庁による桜島の空振計の観測箇所は東郡元、瀬戸、横山、あみだ川の 4 か所である (図 2.3)。

また気象庁が過去の火山観測報を web 公開している [10]。その中でも気象庁が噴火口付近に設置している空振計の値が記載された web ページの情報だけを抽出し、火山噴火が起こった時刻、空振計名と空振計が計測した値についてリストに整理する。今回は機械学習に必要なデータセットを集めるという目的があるためデータ数は多ければ多いほどよい。そのため、空振計ごとに分けたとき、火山噴火の観測回数が最も多く、かつこれまでインフラサウンドの研究が進んでいる桜島の噴火時刻を用いることにする。

2.3.2 訓練データにおけるデータ長の決定方法

気象庁の火山噴火情報に記載されている時刻を基に、噴火に対応する時刻のインフラサウンドデータを参照できる。しかし、火山噴火によるインフラサウンド波形の到達予想時刻が



図 2.3 気象庁による空振計の設置箇所。赤色のピンが気象庁が設置している空振計の設置箇所 [8] であり、青色のピンは桜島御岳である。桜島御岳の経緯度は国土地理院から取得した [9]。

得られたとしても継続時間を知らない限りは終了時刻を定めることができない。本節では特に桜島の火山噴火波形の継続時間、すなわち訓練データのデータ長の定義を定める。

桜島が噴火したときの衝撃波によるインフラサウンドが高知県内の各センサーまで到達するには最も近い宿毛市小筑紫町田ノ浦のセンサー設置個所で 240 km であり、最も遠い東洋町生見のセンサー設置箇所で 400 km のような長距離を伝播する。そのため、対流圏内の水平伝播だけではなく高知県内の上空 (対流圏、成層圏、中間圏および熱圏) においてインフラサウンドが屈折し届く波形を合わせて最大 3 回程度の到達波形が確認できる。これらの到達時間を予測する場合、音速を仮定する必要がある。桜島の噴火による各伝播速度は対流圏の場合約 340 m/s、中間層 (成層圏から中間圏) 屈折の場合は約 311 m/s、熱圏屈折の場合で最も遅く到達する伝播速度は約 263 m/s である [22]。このことから始めに火山噴火のインフラサウンド波形がセンサーに到達するのは対流圏水平伝播によるものだと考えられ、また火山噴火インフラサウンド波形が最後に到達するのは熱圏屈折した波となる。これらを考慮し、対流圏水平伝播、中間層屈折伝播、熱圏屈折伝播を引き起こしたと考えられる火山噴火波形の一例を図 2.4 を示す。黄色、オレンジ、赤色の実線は気象庁による火山噴火時刻から計算された到達時刻であり、黄色破線について気象庁による火山噴火タイムスタンプが分精度のために対流圏水平伝播到達時刻より -30 秒した時刻であり、赤色破線は熱圏屈折伝播到達時刻より +30 秒した時刻である。

ここで、火山噴火によるインフラサウンド波形の継続時間は与えられていないことに注意

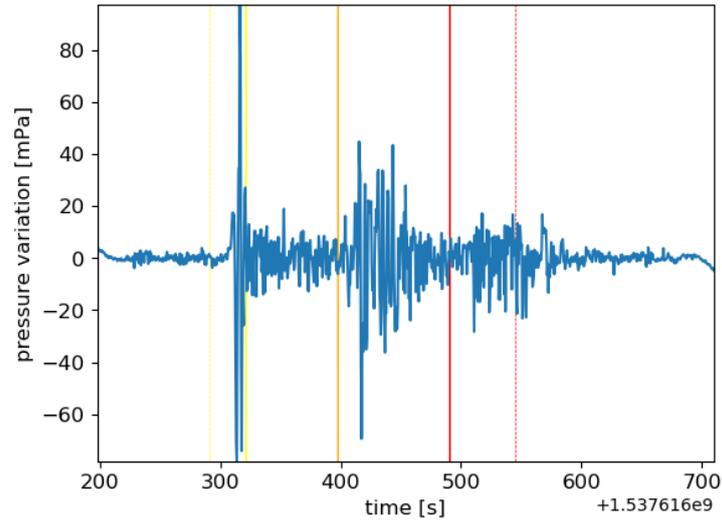


図 2.4 火山噴火によるインフラサウンドが対流圏水平伝播、中間層屈折伝播、熱圏屈折伝播したと考えられる波形。この波形は 2018 年 9 月 22 日 20 時 25 分の桜島の噴火を黒潮町馬荷に設置したセンサーで観測し、低周波数ノイズを除去したグラフ

する。この継続時間を火山噴火によるインフラサウンド波形のスペクトルに含まれる最小周波数から得られる最大周期を 1 周期分とすると、最小周波数を求めるアルゴリズムによって訓練データのデータ長の基準が定まる。火山噴火波形の継続時間を d 、対流圏水平伝播到達時刻および熱圏伝播到達時刻をそれぞれ t_{tropo} , t_{themo} 、噴火の観測時刻誤差を e_t とすると、理論上の訓練データにおけるデータ長は $L = t_{themo} - t_{tropo} + d + 2e_t$ となる。 L について、ある自然数 n に対して $2^n \leq L \leq 2^{n+1}$ を満たすが、FFT(Fast Fourier Transform) による 2 のべき乗の制約および、窓関数により両端部分に関して十分なスペクトル強度が得られないことから、実際のデータ長は $L = 2^{n+1}$ とする。本研究ではデータ長は 1024 と決定した。実際、気象庁による噴火時刻を基にしてリモートセンシングによる火山噴火波形を取り出すためには基準となる時刻が必要である。今回の場合は $(t_{tropo} + t_{themo})/2$ を基準時刻 (中心時刻) として、両端 512 サンプルの位置の幅をデータ長としている。

2.3.3 低周波数ノイズ除去および火山噴火波形の最小周波数を求めるためのアルゴリズム

インフラサウンドセンサーで取得した時系列データの周波数領域において、火山噴火の周波数よりも低周波数側に非常に大きなスペクトル強度を持つノイズ成分が含まれることがある (図 2.7)。低周波数ノイズは定常的ではなく、時刻やセンサーの設置箇所によって低周波数ノイズの種類が異なると考えられ、恣意的ではなくデータに基づく判断から帯域調整が可能なハイパスフィルタ (動的ハイパスフィルタ) を用いる必要がある。ここで便宜上、低周波数ノイズ成分のカットオフ周波数を火山噴火波形の最小周波数であると仮定する。その仮定からその逆数 (周期) は訓練データの決定に必要な火山噴火波形の継続時間と見做すことができる。つまり、低周波数ノイズを除去する際、訓練データにおけるデータ長を求める要素としての最大周期は同時に求められる。インフラサウンドセンサーの観測データに含まれ

る低周波数ノイズ成分の除去にはパワースペクトル強度の平均値 μ と標準偏差 σ を利用した閾値 $\mu + \sigma$ を利用する。本節では火山噴火波形の最小周波数を求めるために用いたアルゴリズムについて説明する。

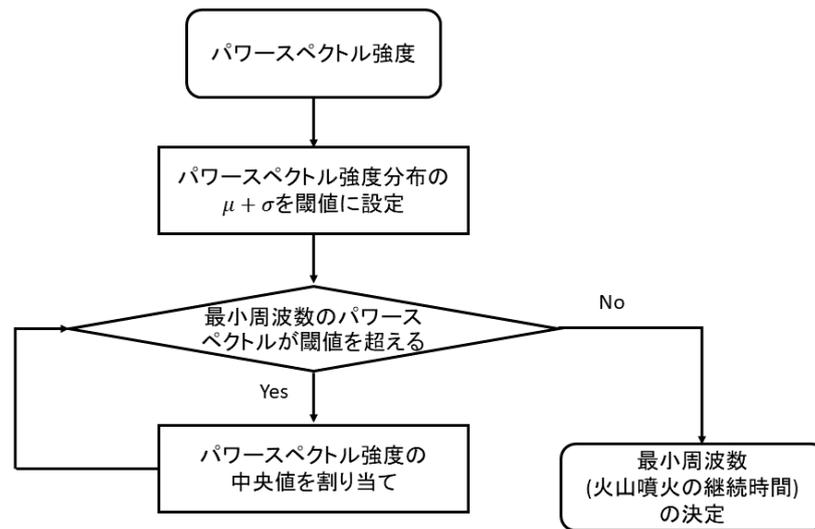


図 2.5 低周波数ノイズ成分の除去および火山噴火継続時間決定方法のフローチャート

図 2.5 において、初めににサンプリングレートを一定にしたデータのパワースペクトルを求める。次にパワースペクトル強度分布の $\mu + \sigma$ を閾値として、周波数分解能ごとに最小周波数のパワースペクトルが閾値を超えているかどうかを判定する。閾値を超えていれば、スペクトルの中央値を求めて割り当てる。閾値を超えなければ、そのデータの最小周波数であり、求めるべき火山噴火波形の最小周波数となる。

ここで、実際に上記のフローチャートを行ったときのデータを 1 例だけ示す。インフラサウンドセンサーの ADXIII INF01LE は 0.001 Hz から 6.25 Hz の帯域を有し、その低周波数側に含まれるノイズを解析するにあたり、サンプリングレートを整えた火山噴火波形を図 2.6 に示す。ただし、図 2.6 の波形は桜島が 2018 年 9 月 22 日 20 時 25 分に噴火し、黒潮町馬荷で観測した波形である。

横軸は UNIX 時間で、縦軸は気圧変動 [mPa] である。300 秒に対してピークが 250 mPa 程度の信号が見られ、400 秒でピークが 100 mPa 程度のピークが確認できる。また長周期で気圧変動が高い波形が確認できる。図 2.6 の波形の振幅スペクトルを図 2.7、パワースペ

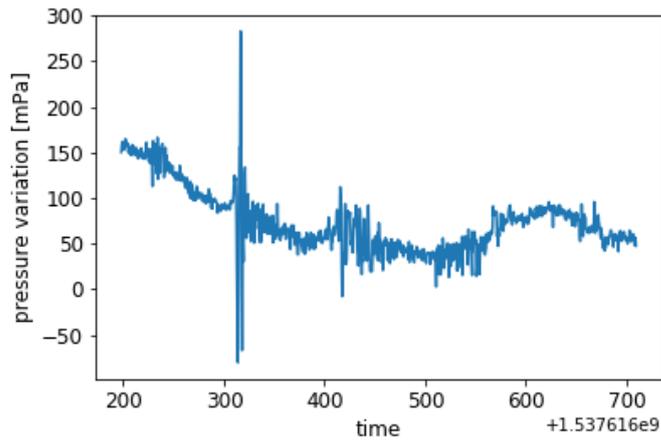


図 2.6 2018 年 9 月 22 日 20 時 25 分に桜島の噴火を黒潮町馬荷で観測した波形

クトルを図 2.8 に示す。

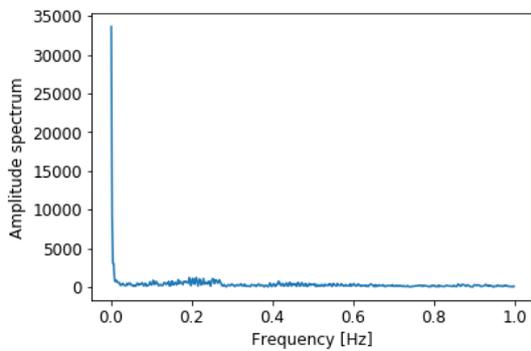


図 2.7 2018 年 9 月 22 日 20 時 25 分に桜島の噴火を黒潮町馬荷で観測した波形の振幅スペクトル

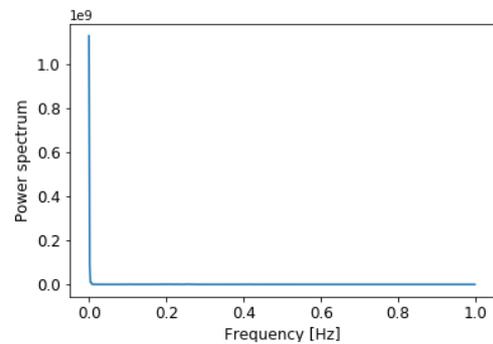


図 2.8 2018 年 9 月 22 日 20 時 25 分に桜島の噴火を黒潮町馬荷で観測した波形のパワースペクトル

図 2.7、図 2.8 の横軸はともにナイキスト周波数の 1 Hz までをプロットしている。周波数分解能である 0.002 Hz に非常に大きなスペクトル強度が表れており、0.002 Hz 以上の周波数帯では振幅スペクトルでは信号はわずかに確認できるが、パワースペクトルでは描画には表示されなくなっている。ここでパワースペクトルに動的ハイパスフィルタをかける。このとき低周波数ノイズを判定するための閾値はデータ長を 1024 と仮定したときのパワースペクトルの分布における平均値 μ と標準偏差 σ を利用する。パワースペクトルのヒストグラムを図 2.9 に示す。

ここで図 2.9 の横軸は得られた波形のパワースペクトルでビン数を 100 としている。また縦軸はヒストグラムにおける度数である。図 2.6 におけるパワースペクトルは 10^9 スケールに分布しているデータをわずかに持つ。そのため、縦軸を対数スケールにして描画を強調

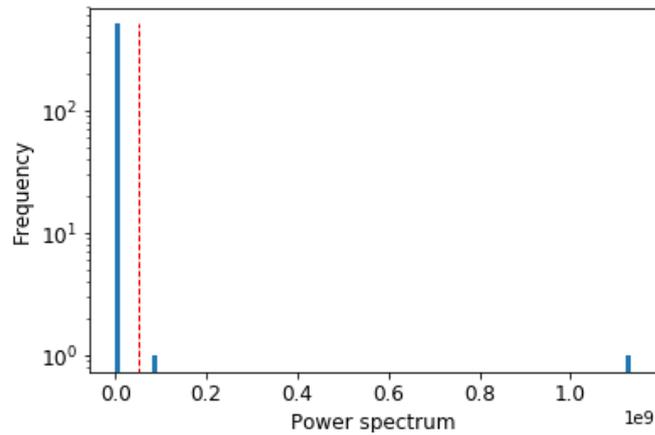


図 2.9 2018 年 9 月 22 日 20 時 25 分に桜島の噴火を黒潮町馬荷で観測した波形のパワースペクトルのナイキスト周波数までのヒストグラム

している。赤破線は今回定義した閾値である。高周波数帯にある信号の度数は低周波数ノイズの度数に対して非常に少ないために平均値は低いパワースペクトルに位置するが、標準偏差は平均値と比較すると大きな値となった。そのために平均値 μ に標準偏差 σ を加えた式 $\mu + \sigma$ を閾値とした。

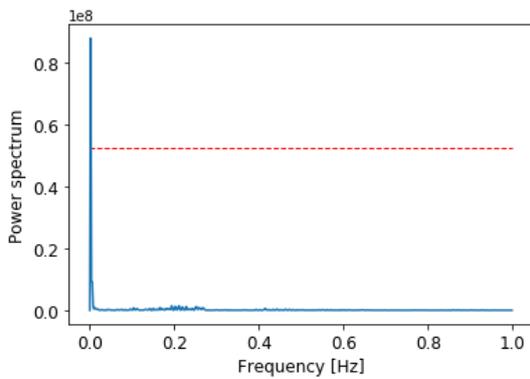


図 2.10 $1 \times$ (周波数分解能) にハイパスフィルタがかかったパワースペクトル

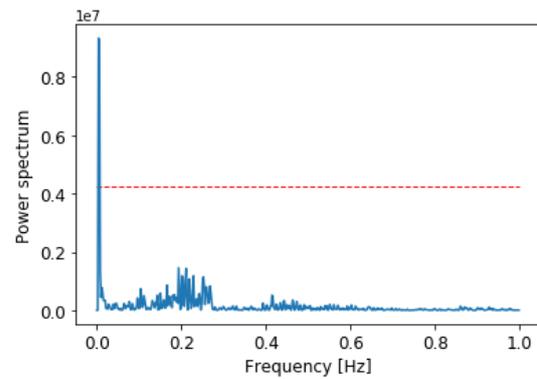


図 2.11 $2 \times$ (周波数分解能) にハイパスフィルタがかかったパワースペクトル

図 2.10 から図 2.20 における赤破線は $\mu + \sigma$ で定まった閾値である。今回の火山噴火波形の場合、周波数分解能 0.002 Hz から 11 回のハイパスフィルタの更新がかかったので、最終的なハイパスフィルタによるカットオフ周波数が 0.0022 Hz に決定する。最終的なハイパスフィルタが定まったときの波形を図 2.21。

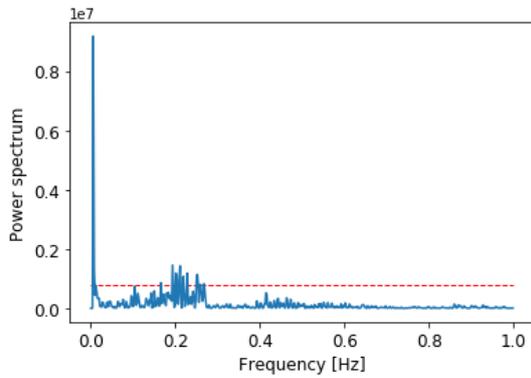


図 2.12 3 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

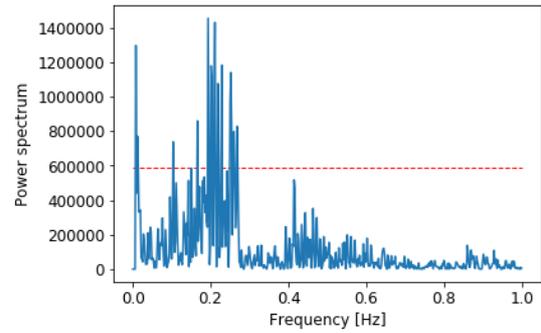


図 2.13 4 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

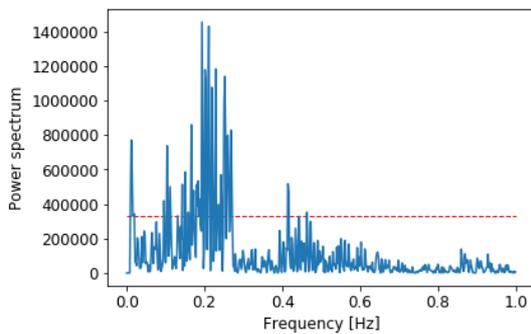


図 2.14 5 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

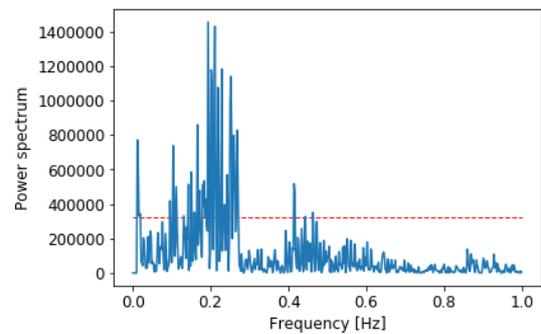


図 2.15 6 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

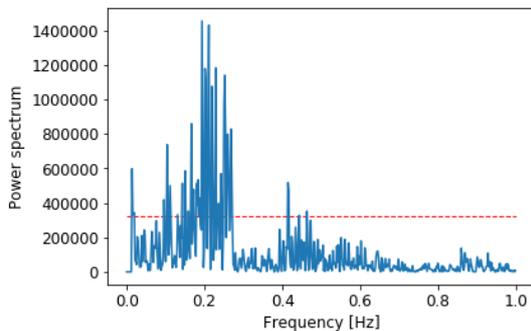


図 2.16 7 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

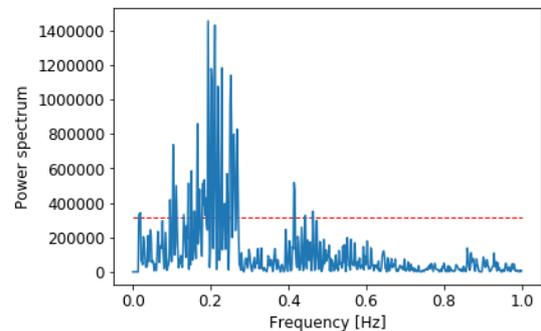


図 2.17 8 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

2.3.4 相互相関関数による他センサーが取得した火山噴火波形の選定

本研究の目的である時系列データをある窓幅でスキャンして火山噴火波形をリアルタイム検知することを想定した場合、センサーの種類や設置箇所によってノイズの特性が異なると考えられることから、センサーごとにデータセットを分ける必要がある。しかし、そのようになるとセンサーの数だけデータセットの数が減ってしまい、個々のデータセットで過学習

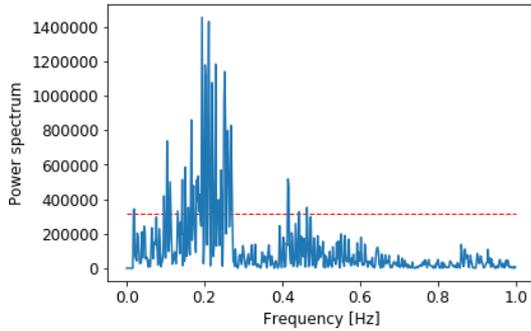


図 2.18 9 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

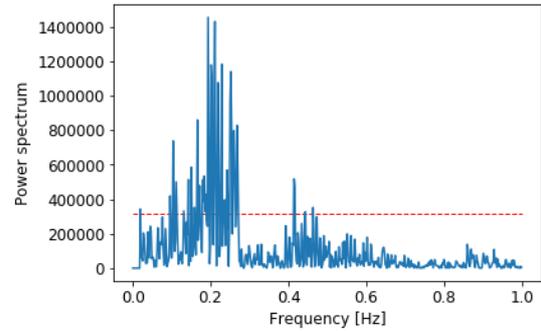


図 2.19 10 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

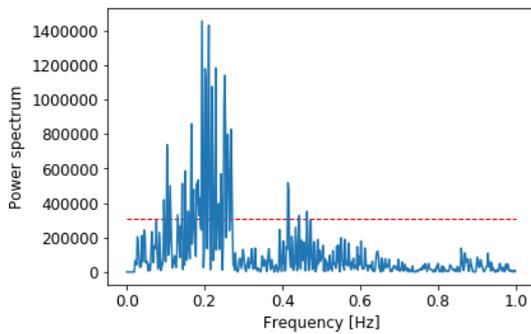


図 2.20 11 × (周波数分解能) にハイパスフィルタがかかったパワースペクトル

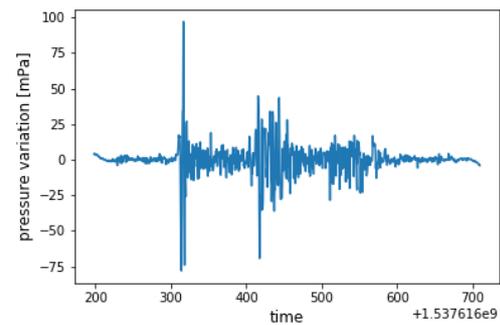


図 2.21 2018年9月22日20時25分に桜島の噴火を黒潮町馬荷で観測した波形に動的ハイパスフィルタをかけたときの波形

に陥る恐れがある。それを避けるために本研究室が複数センサーで面的に同時観測している利点を用いる。具体的には多地点の観測データについて、相互相関処理によって火山噴火波形を検出する手法 [23] において類似度が高い波形であればデータセットとして加えてよいと解釈する。ある1回の火山噴火によるインフラサウンド波形を観測した1台のセンサーに注目したとき、同時観測した別のセンサーの火山噴火波形との相互相関処理を行い、相関が高ければターゲットとなるセンサーのデータセットとして用いてよいとする。相関値を得る手段としてはターゲットのセンサーと他のセンサーとの相互相関関数を用いる。この処理を気象庁が提供している火山噴火時刻に基づいて複数の火山噴火で行う。この処理の流れを図 2.22 に示す。

ここで今回用いた相互相関関数の扱いについて。ある1回の火山噴火について、複数ある各センサー j ($j = 1, \dots, l$) とターゲットとなる1台のセンサー i のそれぞれでデータ長 N の多変量時系列 $\{y_1(j), \dots, y_N(j)\}$ が観測されたとき、平均 $\mu(i)$ 、相互共分散関数 $C_k(i, j)$ および相互相関関数 $R_k(i, j)$ の推定値 $\hat{\mu}(i)$, $\hat{C}_k(i, j)$, $\hat{R}_k(i, j)$ は (5) 式から (7) 式により求

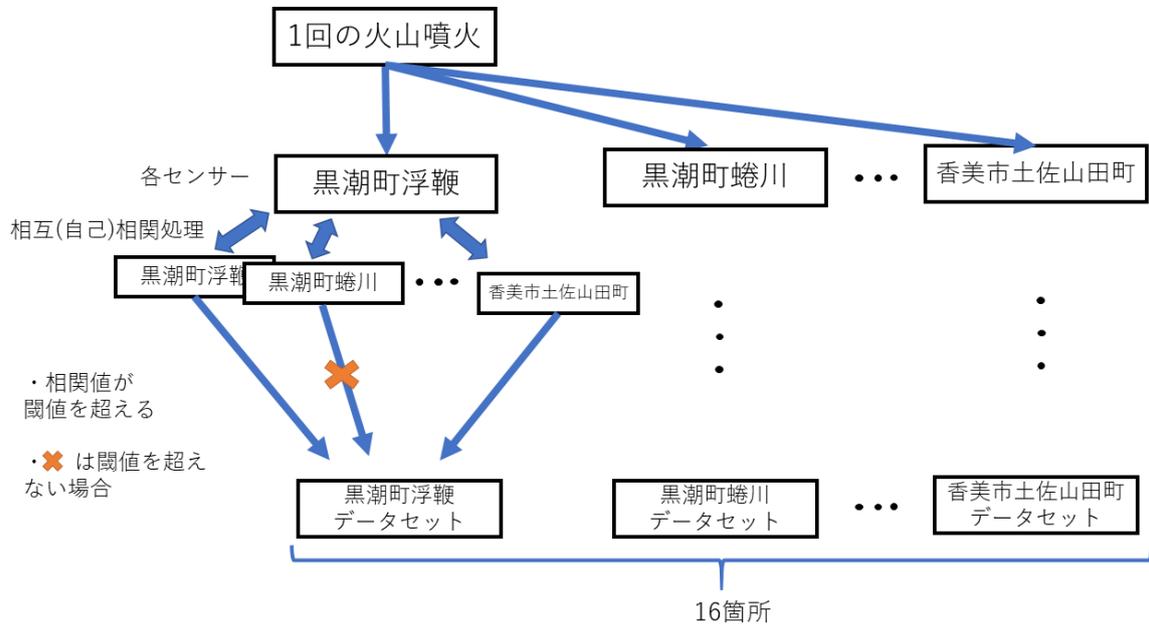


図 2.22 1 回の火山噴火に対して相互相関処理により各センサーのデータセットを増やす処理を複数の火山噴火で行うための概要図

められる [24].

$$\hat{\mu}(i) = \frac{1}{N} \sum_{n=1}^N y_n(i) \quad (5)$$

$$\hat{C}_k(i, j) = \frac{1}{N} \sum_{n=k+1}^N (y_n(i) - \hat{\mu}(i))(y_{n-k}(j) - \hat{\mu}(j)) \quad (6)$$

$$\hat{R}_k(i, j) = \frac{\hat{C}_k(i, j)}{\sqrt{\hat{C}_0(i, i)\hat{C}_0(j, j)}} \quad (7)$$

ただし、 k ($1 \leq k \leq N$) はタイムラグである。これにより、あるセンサーの組 (i, j) における相互相関関数についてデータ長 N 個分の相関値群 $R(i, j) = \{R_1(i, j), \dots, R_N(i, j)\}$ が得られ、タイムラグ k を変数とした一つの時系列グラフと見做すことができる。

ここからは相互相関処理後の各センサーにおけるのデータセット選定のアルゴリズムを説明する。火山噴火が起こっていない場合の各センサーにおけるノイズ成分同士の相互相関関数で得られた相関値の分布を正規分布と仮定する。火山噴火が起こり、各センサーが火山噴火波形をそれぞれ捉えた場合の相関値の分布は、ノイズの散らばり幅を数倍した位置に新たな分布として現れる。この新たな分布を無相関時の正規分布に対する異常値と見做す。これを指標に火山噴火のデータセットとして選定するためには、相関値群の散らばりを用いた閾値の設定が有用である。閾値を設定した場合において、閾値を超えたデータの分布が存在する場合と、相関値群が閾値を超えなかった場合を図 2.23、図 2.24 に示す。

用いた閾値の定義を述べる。ある火山噴火が起きたときのセンサーの組 (i, j) の相関値群の平均値を $R_\mu(i, j) = R_\mu$ 、標準偏差を $R_\sigma(i, j) = R_\sigma$ と両定義したとき、任意定数 t で閾

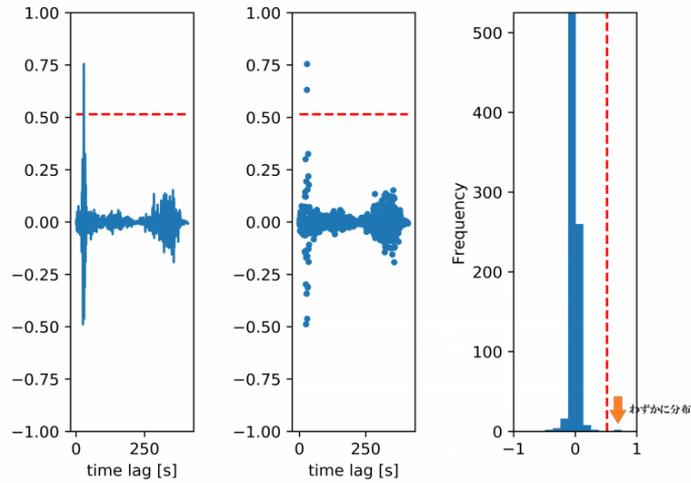


図 2.23 2018 年 5 月 23 日 23 時 58 における桜島火山噴火における黒潮町浮鞭と黒潮町上川口に対する相互相関関数 (左)、散布図でプロットしたグラフ (中央)、相関値群のヒストグラム (右)。赤破線が相関値群の統計値で設定した閾値である。閾値に対して少なくとも 1 点以上が閾値を超えたためインフラサウンドによる火山噴火波形が到達したと判断された。

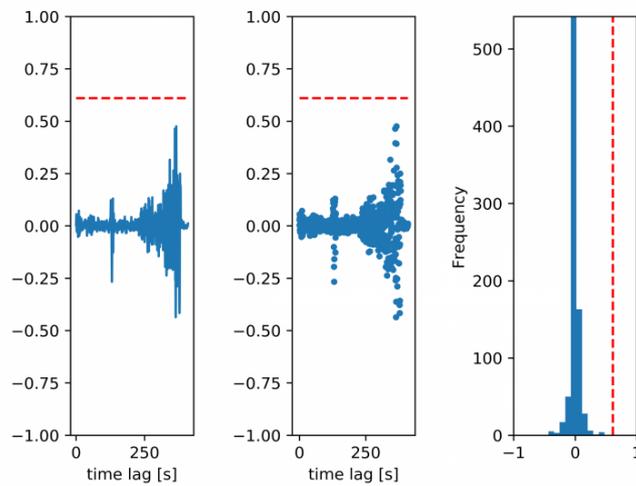


図 2.24 2018 年 5 月 23 日 23 時 58 における桜島火山噴火における黒潮町浮鞭と土佐清水市足摺岬 1 の波形データに対する相互相関関数 (左)、散布図でプロットしたグラフ (中央)、相関値群のヒストグラム (右)。閾値 (赤破線を超えなかったため)、黒潮町浮鞭に対して土佐清水市足摺岬 1 のセンサーには火山噴火波形が到達していないと判断された。

値を変化させるとすると、閾値 T はタイムラグに関わらず $T(t) = R_\mu + tR_\sigma$ となる。この任意定数のことを本論文では閾値倍率 t と呼ぶことにする。このことは信頼係数 α と正規分布の標準誤差 $\frac{\sigma}{\sqrt{N}}$ を考えたとき、正規分布の上側 $100(1 - \alpha)\%$ 点を $z_{\frac{1-\alpha}{2}}$ とすると、このときの閾値倍率 t は $z_{\frac{1-\alpha}{2}} \times \sqrt{N}$ に相当する。ここで火山噴火データセットとして扱ってよい条件を述べる。あるタイムラグ k のときの相関値を $R(k)$ とする。 $T(t) < R(k)$ となる k が存在するとき、ターゲットの 1 台のセンサー i に対する他センサー j ($i \neq j$) のデータはターゲット 1 台のセンサーのデータセットとして使用可能であると判断する。このとき、ターゲットのセンサー i において同じ組 (i, i) が選ばれたときは自己相関関数となり、タイムラグ $k = 0$ のときは相関値 $R(0)$ は必ず 1 となる。実際の判定では次の 2 点に留意する必要がある。(i) 標準偏差 R_σ が十分に大きい場合、閾値 $T(t)$ は 1 を超えてしまう場合があり、ターゲットのセンサーデータを破棄すると判定してしまう。他センサーとの比較によってデータセットの選定を決定したいため、便宜上、自己相関関数の相関値群 $R(i, i)$ には常に $T(t) < R(k)$ となる k があると判定する。(ii) ある火山噴火情報に対する相互相関関数 $R(i, j)$ がすべて閾値 $T(t)$ を超えない場合はターゲットのセンサーのみとなる。そのような場合は (i) のルールから、ターゲットのセンサーのみのデータセットしか存在しないという判断となる。その場合は、この火山噴火情報に対するデータはすべて棄却する。本研究では、(i), (ii) 以外の場合のみを各センサーに対するデータを扱ってよいと判定する。数式に表す 8 式のようなになる。

$$J(i, j) = \begin{cases} 1 & (T(t) < R(k) \text{ となる } k \text{ が存在する または } i = j) \\ 0 & (\text{それ以外}) \end{cases} \quad (8)$$

このときセンサーの個数 l に対する $l \times l$ 行列 $R = [J(i, j)]$ における行ベクトル $R(i) = (R(i, 1), \dots, R(i, l))$ について、 $\sum_{j=1}^l R(i, j) \geq 2$ を満たすとき、 $J(i, j) = 1$ に対応するセンサー j とターゲットのセンサー i のデータを火山噴火データセットとする。(i) の条件は行列 R の対角成分がすべて 1 であること、すなわち $J(i, i) = 1$ に対応し、(ii) の条件は $\sum_{j=1}^l R(i, j) = 1$ に対応している。

この方法は 1 回の火山噴火において、16 台のセンサーでは高々 16 個しか収集できない。複数の火山噴火イベントのデータもについてこの方法を適用することで多くの火山噴火波形のデータセットを集めることができる。

ここで閾値倍率 t を変化させると、データセット数も変化する。閾値倍率 t を高めると高いノイズレベルの波形データが破棄されることからデータセット候補数 $N_s(t)$ ($t \geq 0$) と呼ぶことにする。閾値倍率 t を説明変数としたデータセット候補数の変化の実験を 3.3 節で述べる。閾値倍率 t を上げると、相関値群 $R(i, j)$ の散らばりは小さいものの、あるタイムラグ k のときには高い相関値が得られるようなグラフのみがデータセットとして判定され、データセットの質 (同質性) は高まると考えられる。このことは、閾値 $T(t)$ が低いもしくは設定しないと $T(t)$ を非常に高くしたときの分類器による性能評価値の比較をすることで検証できる。また標準偏差 R_σ が小さい相関値群に対しては、閾値倍率 t を少し上げ

でもあるタイムラグ k における相関値 $R(k)$ は閾値 $T(t)$ を超えたままだと考えられるので、そこを閾値倍率 t の決定とすべきである。その閾値倍率 t のところで各観測点ごとにデータセット数を定めて、分類器による性能評価値を計測する実験を 3.5.1 項で説明する。

2.3.5 火山噴火波形以外のデータセット収集条件

これまでは火山噴火波形となる p^* (positive) のデータの選定方法についてを述べた。分類のためにははそれ以外の波形 n^* (negative) の種類の分布を定義する必要がある。 n^* のデータを構成するにあたり、以下の 3 条件を考慮した。

- (i) n^* のデータは時刻によって取得することで、データ取得の再現性が加味されている
- (ii) p^* ではターゲットのセンサーだけでなく、相互相関処理によって他センサーのデータが入っているため、それに対応するノイズデータを n^* のデータに含める
- (iii) 火山噴火だけではなく、今回定義した訓練データにおけるデータ長と同等の継続時間がある大規模自然災害のデータを含める

ここで、(iii) に対しては例えば地震、雷が当てはまるが、火山噴火波形の分析と同様に、到達時刻が正確に導出できなければいけないため、本研究のデータセットの構成には入れておらず今後の課題とした。条件 (i),(ii) に対しては次の手順でデータセットを集めることで解決することができた。

1. 火山噴火波形を収集する際、火山噴火の発生時刻から 30 分間のデータを抽出する。
ここで火山噴火波形の到来予測は桜島から高知県内のセンサー群までの距離を仮定した場合、この 30 分間に含まれている
2. p^* のデータ長の最初時刻から同じデータ長だけ時刻を遡ったデータを切り取り、 n^* のデータとする
3. 2. でデータ長が足りない場合は、 p^* のデータのデータ長の最後の時刻から同じデータ長だけ切り取り、 n^* のデータにする
4. 相互相関処理で p^* のデータとして選定したセンサーに対応して、同数の n^* のデータも抽出する

手順 2. 3. では手順 1. のデータ取得方法に依存しているだけであり、条件 (i) を満たしている。また条件 (ii) は手順 4. で満たすことができる。以上により n^* のデータセットを構成した。この取得方法を次のフローチャート (図 2.25) にまとめた。

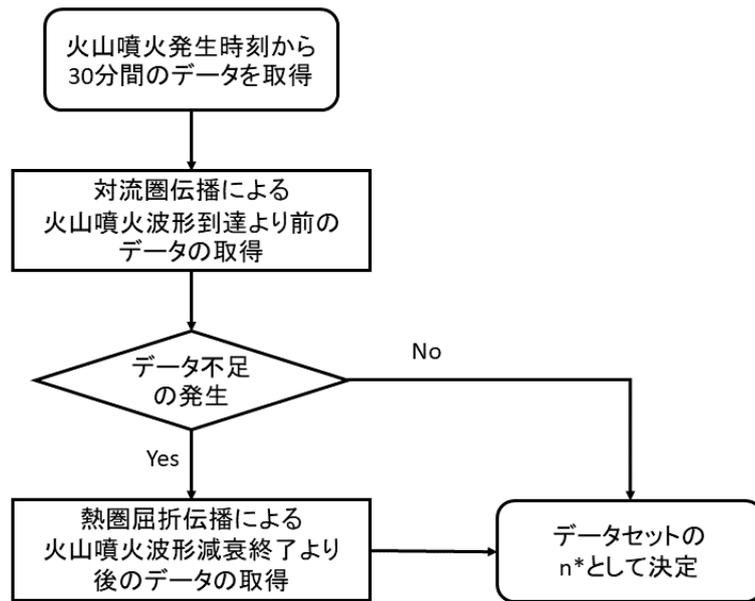


図 2.25 n^* のデータの取得方法のフローチャート

2.4 ケプストラム分析と MFCC による特徴量抽出

2.4.1 ケプストラム分析

ケプストラム分析は信号波形の対数パワースペクトルを取ったうえで、さらにフーリエ変換することで得られる (図 2.26)。ここで、ある信号 $x(t)$ が偶関数である場合、フーリエ変換 $\mathcal{F}[x(t)]$ をした後、さらにフーリエ変換した信号 $\mathcal{F}[\mathcal{F}[x(t)]]$ の領域は時間であることが示される [25]。しかし、ケプストラムでは 1 度目のフーリエ変換の後に対数をとっているため、元の時間領域とは異なる。この時間領域のことをケフレンシー (quefrensy) という。

ケプストラム分析は畳み込み積分で表される複数の信号を、和の形で扱える領域に変換する処理の一つである。音声工学において声道成分のスペクトル $V(f)$ と基本周波数に相当するスペクトル $S(f)$ で構成される音声信号のスペクトル $X(f) = V(f)S(f)$ は、ケプストラム領域において対数のフーリエ変換 \mathcal{F} を施した足し算の形

$$\mathcal{F}[\log |X(f)|] = \mathcal{F}[\log |V(f)|] + \mathcal{F}[\log |S(f)|] \quad (9)$$

として分解することができる。声道成分となるケプストラムの低次係数はパワースペクトル包絡となり、高次係数は基本周波数の抽出ができる。特にパワースペクトル包絡を抽出することは声道特性の抽出に相当し、その極大であるホルマントの分布を得ることは母音の発音など発音認識に用いられる。ケプストラム分析だけではフレーム長だけデータがあるため、実用的ではない。そのため、通常フィルタバンクを用いてデータを低次元に圧縮させる方法

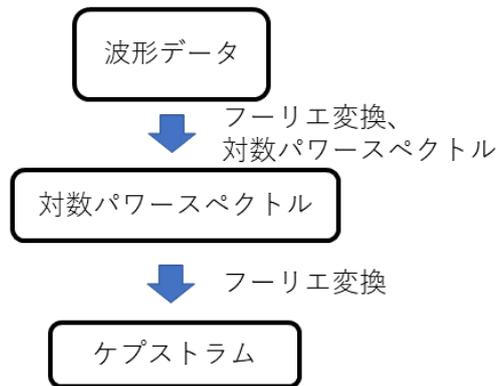


図 2.26 ケプストラム分析のアルゴリズム

がとられる。ここでフィルタバンクとは通過帯域が適当に設定されたフィルタをいくつか並列に並べて音声信号の周波数成分をフィルタの出力の組として得られるものとされている [26]。

2.4.2 MFCC による音声の特徴量抽出

ケプストラム領域における周波数軸の尺度は線形のままである。ここで人間の聴感覚特性が周波数に対してほぼ対数的であることを利用したメル尺度を用いて、分布を非線形化する。メル尺度は音の高さの心理尺度を求めるために周波数の異なる純音を与え、感覚が等しくなるように、その間を 2 分割する周波数を求める実験の曲線であるとされている [27]。その実験におけるメル尺度 m の近似式における変種は多いが、今回は次の式を用いる [28]。

$$m = m_0 \log \left(1 + \frac{f}{f_0} \right) \quad (10)$$

ここで f_0 は周波数パラメータである。 $f_0 = 700$ Hz としたとき、 m_0 は 1 kHz で 40 ホンの純音の高さを $m = 1000$ メルと定めるという制約より (10) 式から $m_0 = 1125$ が導かれる。メル尺度で設計したフィルタバンクをメルフィルタバンクと呼ぶ。メルフィルタバンクを用いることで低周波数帯域では感度を高くし、高周波数帯域では荒くスペクトルをとることができ、高周波数側のスペクトルの情報を圧縮することができる。フィルタバンクで得られたときのフィルタ数のことをチャンネル数と呼ぶ。ナイキスト周波数が 8 kHz の場合のメルフィルタバンクを図 2.27 に示す。横軸は周波数であり、縦軸は透過率を表す。

さらにメルフィルタバンクで指定したパワースペクトルの標本値に対して、離散コサイン変換した係数の組はメル周波数ケプストラム係数 (以下、MFCC) と呼ばれており、音声認識の特徴量として直接用いることが多い。MFCC のアルゴリズムを図 2.28 に示す。

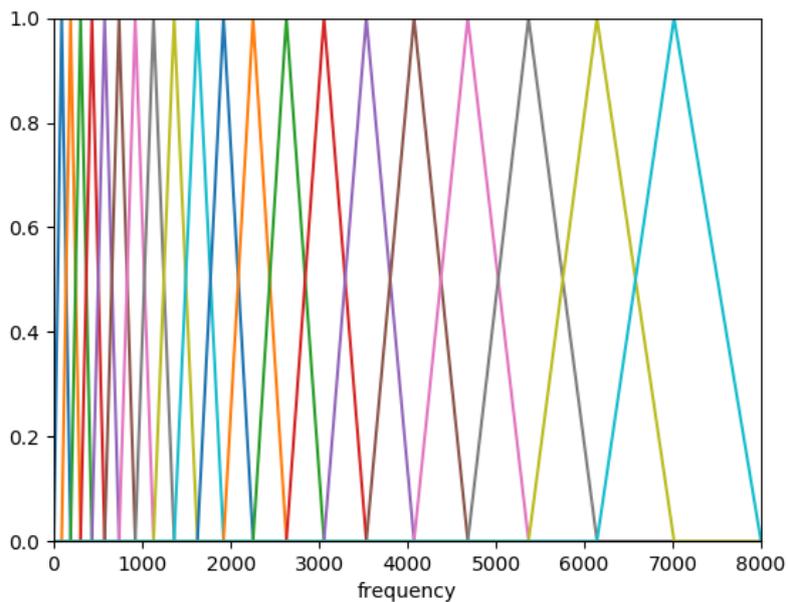


図 2.27 ナイキスト周波数が 8 kHz でチャンネル数が 20 の場合のメルフィルタバンク

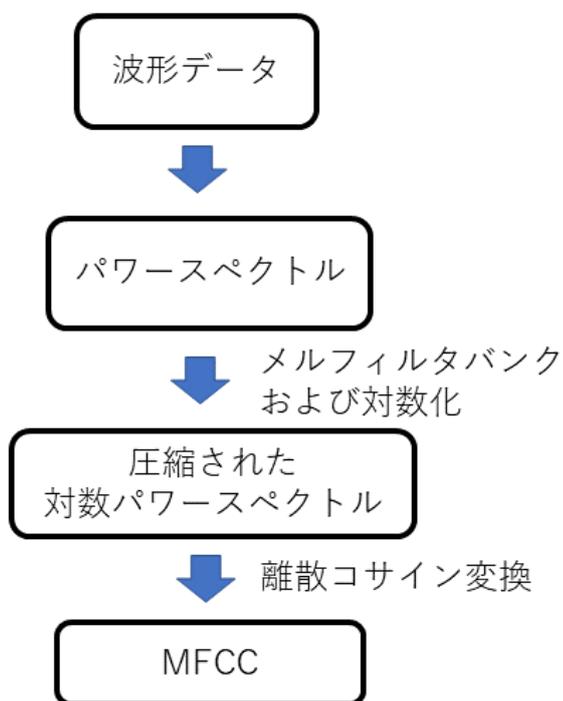


図 2.28 MFCC のアルゴリズム

離散コサイン変換はケプストラム分析における 2 回目のフーリエ変換に相当する。しかし、対数パワースペクトルをメルフィルタバンクで圧縮しただけでは各係数は相関を持っているため、無相関化するために離散コサイン変換を用いるとされている [29]。ここで通例であれば、メルフィルタバンクで 20 次元程度にして、さらに離散コサイン変換した係数を低次から 12 次元だけ抽出する [30]。しかし、一番低次の項は直流成分となる。この項は音声認識では音素の違いを識別するために、この項を抜き取るとされる [31]。離散コサイン変換で得られた 20 個の項 c_0, \dots, c_{19} において、12 次元だけ抽出する場合は直流成分 c_0 を除いた c_1, \dots, c_{12} が今回用いる特徴量となる。

2.4.3 MFCC による動的に変化する場合におけるデルタケプストラムによる特徴量抽出

MFCC はスペクトルの各時刻の静的な特徴量であったが、スペクトルが時間方向に動的に変化する場合の特徴量も重要であるとされる [32]。その場合は隣接したフレーム間でケプストラムの差分を特徴量とする [30]。この方法をデルタケプストラム (Δ ケプストラム) という。具体的な方法として、特に MFCC のではあるフレーム t について、前後のフレーム数を K 、 n 次元目の MFCC を $c_n(t)$ としたときのデルタケプストラム Δc_n は次のように表される。

$$\Delta c_n = \frac{\sum_{k=-K}^K k c_n(t+k)}{\sum_{k=-K}^K k^2} \quad (11)$$

すなわち、MFCC の各次元に対して時間変化の回帰係数を特徴量としている [33]。また (11) 式から MFCC とデルタケプストラムは同じ次元数を持っているため、特徴量の抽出に MFCC とデルタケプストラムを併用しても 2 倍の次元数程度となる。

2.4.4 特徴ベクトルの構成

今回用いるのは MFCC とデルタケプストラムの 2 種類であるため、特徴量として用いられるパターンは、(i) ある訓練データ全体を窓幅とした MFCC のみ、(ii) デルタケプストラムのみ、(iii) MFCC とデルタケプストラムである。(i) と (ii) は同じ次元数であるが、(ii) は特徴量ベクトルが (i)、(ii) に対して 2 倍のデータ長となる。一般に MFCC にデルタケプストラムの特徴量を加えると性能が上昇する [30]。その確認のために (i)-(iii) を 3.4.1 項および 3.5.2 項で検証する。

2.5 可聴音データセットとの対照実験

2.5.1 Speech Commands Dataset について

MFCC は人間の聴感覚を基準にしているため、インフラサウンドに適しているかは不明である。また MFCC およびデルタケプストラムは自作アルゴリズムであるため、致命的なバグがあった場合は性能指標が正しく現れないと考えられる。これらの理由からインフラ

サウンドデータに対して、可聴音データセットを対照実験として用いることにする。今回対照実験として用いたのは Speech Commands Dataset(©TensorFlow team, AIY team, 2017) である。Speech Commands Dataset はディープラーニングを使った音声認識の開発をする場合においてシンプルなタスクをするためのデータセットがないことを背景に、無料で広く利用されるように TensorFlow チームと AIY チームが作成した音声データセットである。Speech Commands Dataset に含まれている音声データは 30 種類の短い英単語を発音した 16 kHz でサンプリングされた長さ 1 秒の音声データがあり、合計で約 65000 個の音声データが含まれている。その音声データは AIY が作成したウェブサイトを通して一般の人が発音した英単語となっている [37]。

2.5.2 火山噴火波形データセットに合わせたデータセットの構成方法

本研究の目的はインフラサウンドにおける火山噴火波形とノイズ波形を 2 クラス分類したときの性能評価値を求めることである。そのため、Speech Commands Dataset もある一つの英単語に対する他の英単語のデータセットの構成にする必要がある。具体的には 30 単語ごとに訓練データおよびテストデータを作成することができ、一つの単語ごとに約 1600 から約 2500 のデータが用意されている。ここで選んだ英単語に対して他の英単語のデータセットは 29 単語から選ぶことができるが不均衡データにならないようにする必要がある。そのために対象のデータ数に対して、他単語データは一様分布による非復元抽出によってデータ数を絞り込み、英単語における p^* データと n^* データの比を 1 : 1 にする。この処理を図 2.29 に示す。ここでラベル 1 は p^* データ、ラベル 0 は n^* データのことを表す。

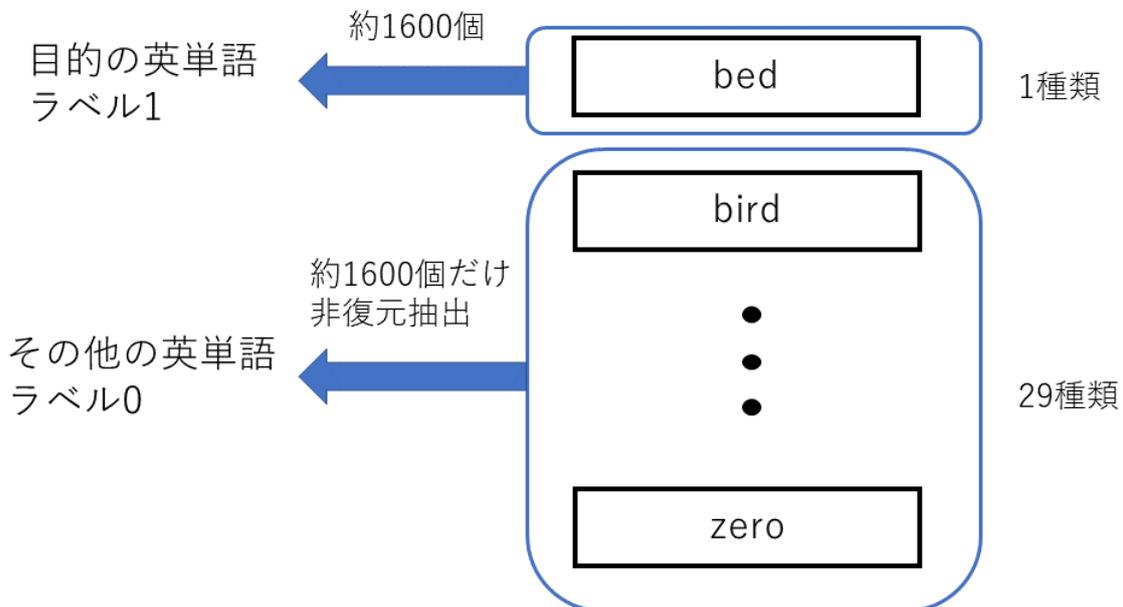


図 2.29 ある英単語と他の英単語を不均衡データにしないための概要図

2.5.3 Speech Commands Dataset と火山噴火波形のパラメータの比較

波形データを MFCC とデルタケプストラムに変換するときいくつかパラメータが必要である。波形データのデータ長だけ MFCC をする場合、データ長と MFCC による周波数パラメータ、デルタケプストラムではフレーム長と窓幅をシフトさせるときのオーバーラップ率が必要である。ここで Speech Commands Dataset と火山噴火波形について、音声処理に関するパラメータおよび MFCC とデルタケプストラムで設定するパラメータの比較を表 2.5.3 に示す。

表 2.2 Speech Commands Dataset と火山噴火波形における音声処理および MFCC とデルタケプストラムのパラメータの比較

	Speech Commands Dataset	火山噴火波形
データ長	16000 (1 秒)	1024 (約 8.5 分)
サンプリングレート [Hz]	16000	2
周波数パラメータ [Hz]	700	0.4
フレーム長	2048 (約 0.13 秒)	256 (約 2.1 分)
オーバーラップ率	0.75	0.75

また一般的にフーリエ変換をする前に波形データには窓関数がかかる必要があるが、今回はハミング窓を使用した。

3 実験結果

3.1 気象庁から取得される火山噴火情報数

気象庁が桜島の火山噴火をターゲットに空振計を設置している場所は、火口周辺において東郡元、横山、瀬戸 2、あみだ川の 4 か所であり、火山観測報は 2014 年 1 月から web 公開されている。そのため火山噴火時刻を取得するためには 1 か所に絞る必要がある。機械学習における性能評価値がデータセット数に依存することを考慮した場合は、データセット数が十分に取得できる観測点を選ぶべきである。今回は 2019 年 11 月 3 日までの火山観測報の中に含まれる桜島の噴火時刻と空振計名と空振値について、4 か所のうち空振値が最も記録されていた観測箇所を用いる。2019 年 11 月 3 日時点に取得した火山噴火時刻および空振値が取得できた個数および、空振値の統計値を表 3.1 に示す。

表 3.1 気象庁による桜島の火山噴火観測地点における空振値の統計値

観測点	記録回数	空振値 [Pa]
東郡元	1515	10.1±0.3
瀬戸 2	1340	40.9±1.0
横山	309	20.1±0.8
あみだ川	23	18.6±2.3

表 3.1 より空振値が最も多く記録されていたのは東郡元観測点であるため、同観測点の火山噴火時刻を利用することにした。

3.2 火山噴火波形の継続時間および訓練データのデータ長の決定

訓練データにおけるデータ長の決定にはインフラサウンドとして取得した火山噴火波形の継続時間が必要である。今回は継続時間 d を火山噴火の最大周期 $T_{max} = 1/f_{min}$ とする。ただし、 f_{min} は火山噴火波形データから求められる最小周波数である。インフラサウンドセンサーの定点観測が開始された 2016 年 8 月 13 日から 2018 年 12 月 31 日までにおいて、東郡元観測地点の空振値が 10 Pa を超えている火山噴火は 27 例であった。そのうち火山噴火波形が目視で入っているとされる時系列データにおいて、対流圏伝播、中間層伝播と熱圏伝播に分けたときのサンプルサイズを N 、火山噴火の最小周波数 f_{min} および最大周期 T_{max} の統計値を表 3.2 に示す。

表 3.2 から火山噴火波形の継続時間は全体の $d = T_{max} = 22$ 秒 とする。誤差を考慮する必要はあるが、FFT における 2 の累乗の制約によって冗長性を加えることにより問題にはならなくなる。

表 3.2 火山噴火波形の最小周波数と最大周期の統計値

記号	サンプルサイズ N	最小周波数 [Hz] f_{min}	最大周期 [s] T_{max}
対流圏水平伝播	56	0.047±0.005	21±2
中間層伝播	24	0.047±0.006	21±3
熱圏伝播	7	0.042±0.008	24±4
全体	87	0.046±0.004	22±2

3.3 相互相関関数を用いて選定された各センサーのデータセット数

1 回の火山噴火に対してリモートセンシングによる火山噴火波形は最大でインフラサウンドセンサーの設置数だけ抽出できる。今回の場合はインフラサウンドセンサーは高知県の香美市土佐山田町宮ノ口、黒潮町浮鞭、黒潮町蜷川、黒潮町上川口、黒潮町出口、黒潮町馬荷、宿毛市小筑紫町田ノ浦、土佐清水市足摺岬 1、土佐清水市足摺岬 2、土佐清水市足摺岬 3、土佐市宇佐町宇佐、高知市春野町芳原、南国市物部、安芸市西浜、東洋町生見、室戸市室戸岬町の計 16 か所となる。1 か所で観測された火山噴火波形とその他の地点で観測された火山噴火波形に対して相互相関処理を行った (図 2.22)。この操作をインフラサウンドセンサーが初めに設置された 2016 年 8 月 13 日から 2018 年 12 月 31 日までに観測された計 27 回の火山噴火に対して行った。ここで、各センサーのデータに対して相関値の閾値倍率 t を説明変数としたときの各センサーにおけるデータセット数の変化をプロットすることができる。このデータセット数は閾値倍率 t によってデータ欠損と高いノイズレベルによって破棄されてしまうことからデータセット候補数 N_s と呼ぶことにする。今回は閾値倍率 t を 0 から 15 まで 0.5 ごとに計算した。このときの各観測点における閾値倍率 t に対するデータセット候補数 N_s の変化を図 3.1 から図 3.16 に示す。

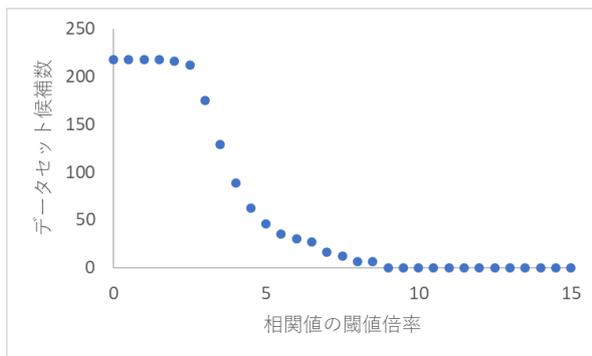


図 3.1 香美市土佐山田町宮ノ口における相関値の閾値倍率に対するデータセット候補数の変化

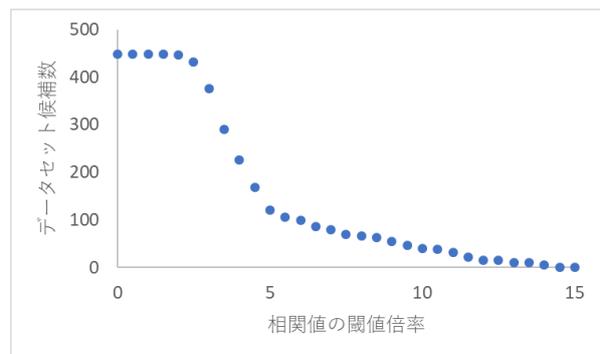


図 3.2 黒潮町浮鞭における相関値の閾値倍率に対するデータセット候補数の変化

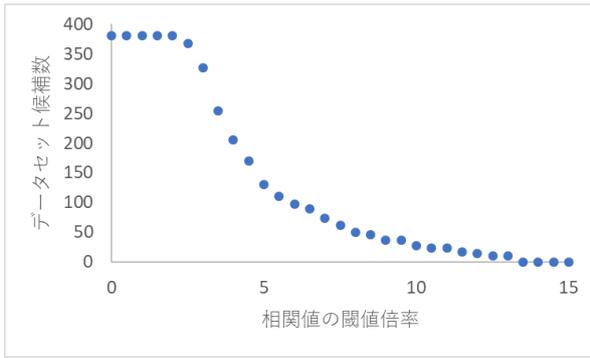


図 3.3 黒潮町蜷川における相関値の閾値倍率に対するデータセット候補数の変化

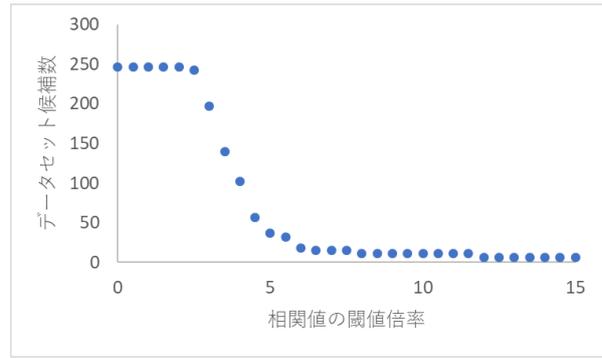


図 3.4 黒潮町上川口における相関値の閾値倍率に対するデータセット候補数の変化

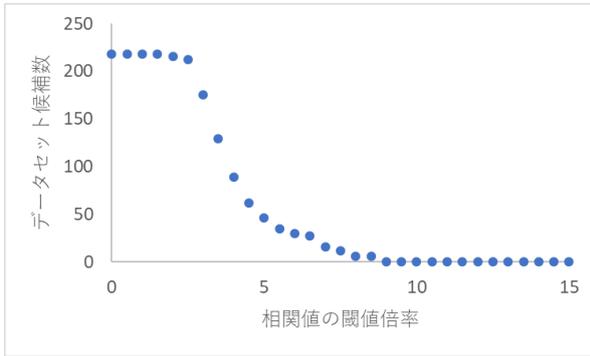


図 3.5 黒潮町出口における相関値の閾値倍率に対するデータセット候補数の変化

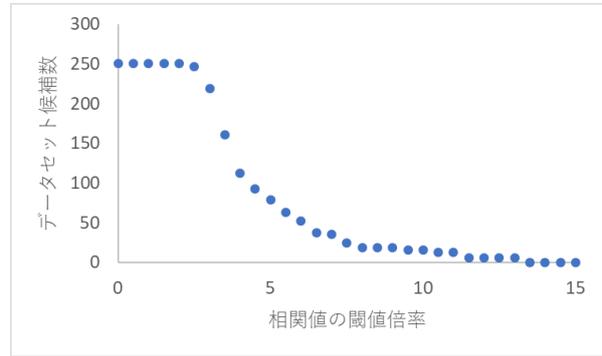


図 3.6 黒潮町馬荷における相関値の閾値倍率に対するデータセット候補数の変化

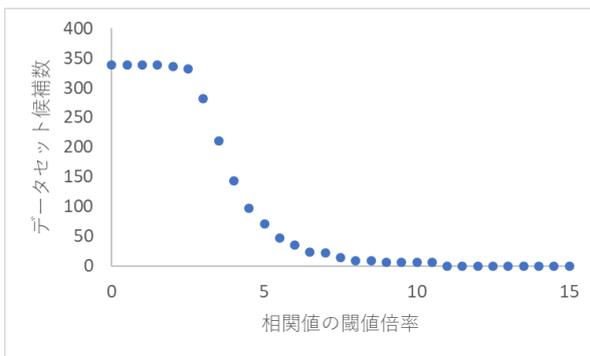


図 3.7 宿毛市小筑紫町田ノ浦における相関値の閾値倍率に対するデータセット候補数の変化

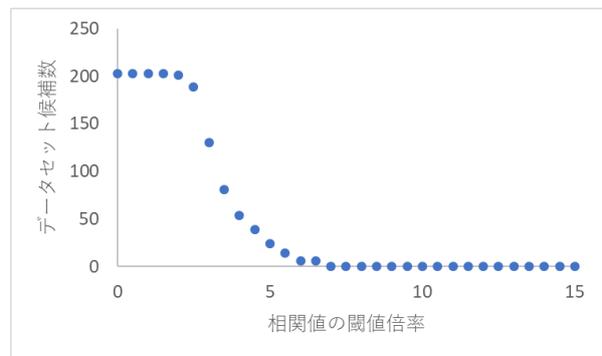


図 3.8 土佐清水市足摺岬 1 における相関値の閾値倍率に対するデータセット候補数の変化

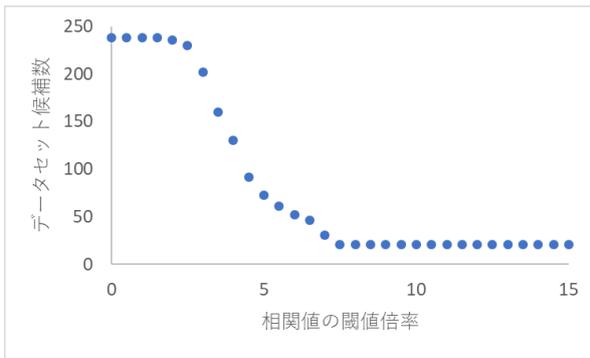


図 3.9 土佐清水市足摺岬 2 における相関値の閾値倍率に対するデータセット候補数の変化

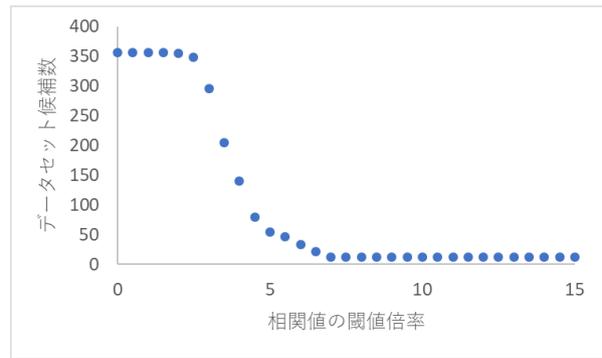


図 3.10 土佐清水市足摺岬 3 における相関値の閾値倍率に対するデータセット候補数の変化

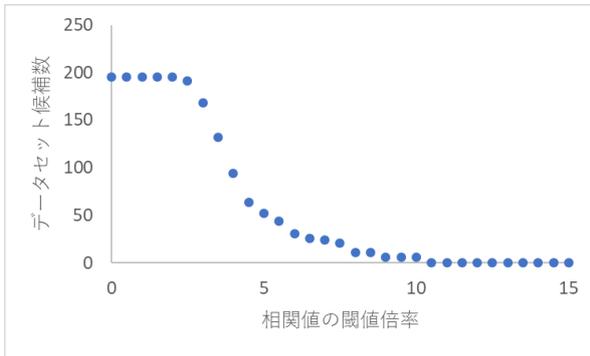


図 3.11 土佐市宇佐町宇佐における相関値の閾値倍率に対するデータセット候補数の変化

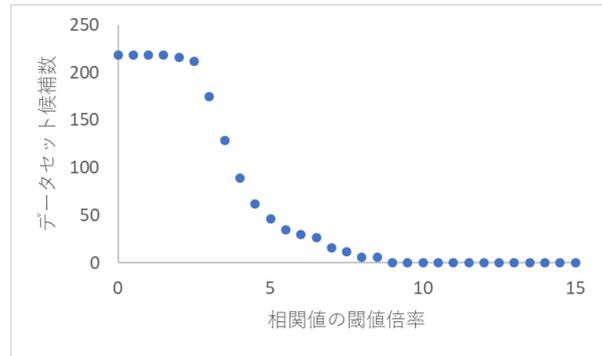


図 3.12 高知市春野町芳原における相関値の閾値倍率に対するデータセット候補数の変化

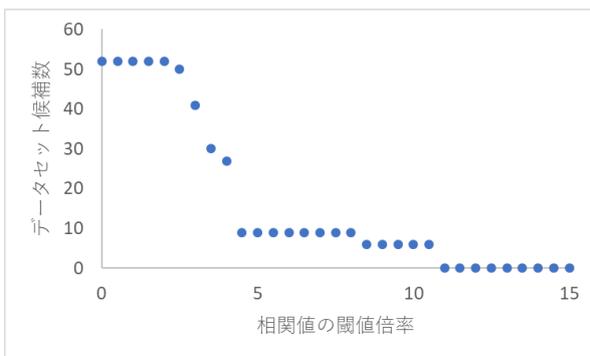


図 3.13 南国市物部における相関値の閾値倍率に対するデータセット候補数の変化

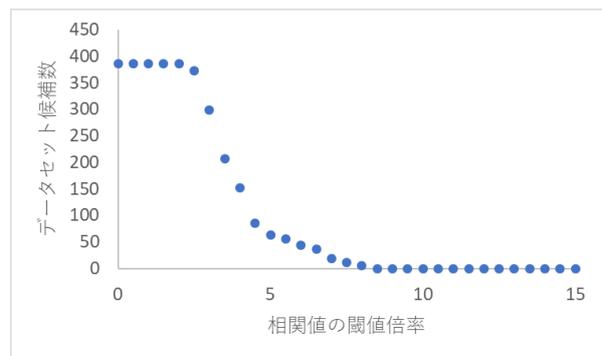


図 3.14 安芸市西浜における相関値の閾値倍率に対するデータセット候補数の変化

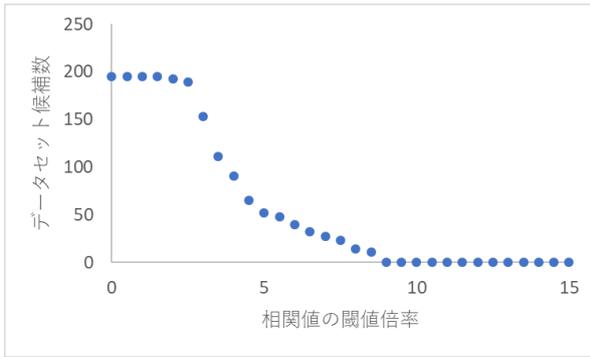


図 3.15 東洋町生見における相関値の閾値倍率に対するデータセット候補数の変化

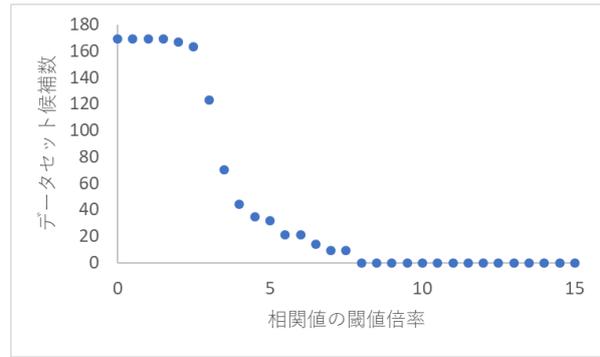


図 3.16 室戸市室戸岬町における相関値の閾値倍率に対するデータセット候補数の変化

どのグラフにおいても閾値倍率 t が約 3 になると、データセット候補数 N_s が急激に落ちていく様子が分かる。その後、閾値倍率 t が約 5 になるとデータセット候補数の減少が緩やかになる。顕著なのは香美市土佐山田町宮ノ口 (図 3.1)、黒潮町出口 (図 3.5) である。その後はこのスケールであれば線形で減少する、もしくは階段状に減少している。閾値倍率が 15 だと土佐清水市足摺岬 2 (図 3.9) と土佐清水市足摺岬 3 (図 3.10) 以外は $N_s = 0$ となった。

3.4 Speech Commands Dataset における特徴量抽出および分類

3.4.1 MFCC またはデルタケプストラムを特徴量とした性能評価値

各英単語について、MFCC、デルタケプストラム、ならびに MFCC とデルタケプストラムそれぞれの組み合わせを特徴ベクトルとして線形 SVM に入力したときの正解率を表 3.4.1 に示す。ただし、デルタケプストラムは Δ と表記する。ここではデータセット数に関しては考慮せず、判別したい p^* のデータの英単語に対して提供されている音声データをすべて使用した。 n^* のデータは p^* でない英単語を p^* のデータと同じデータ数だけ一様分布で非復元抽出した (図 2.29)。そのため、 p^* と n^* のデータ数の比は $P : N = 1 : 1$ となっている。

表 3.3 SVM による単語と特徴量ごとの正解率

英単語	データセット数	正解率		
		MFCC	Δ	MFCC+ Δ
bed	3426	0.83±0.01	0.68±0.02	0.84±0.01
bird	3461	0.87±0.02	0.79±0.01	0.87±0.02
cat	3466	0.85±0.02	0.71±0.02	0.85±0.02
dog	3492	0.77±0.03	0.70±0.01	0.78±0.02
down	4718	0.64±0.03	0.64±0.03	0.70±0.02
eight	4704	0.89±0.02	0.79±0.03	0.89±0.02
five	4714	0.71±0.04	0.63±0.01	0.73±0.04
four	4744	0.84±0.02	0.75±0.02	0.85±0.02
go	4744	0.72±0.03	0.68±0.03	0.74±0.02
happy	3484	0.76±0.02	0.74±0.02	0.80±0.02
house	3500	0.81±0.02	0.82±0.02	0.84±0.02
left	4706	0.68±0.02	0.66±0.02	0.72±0.02
marvin	3491	0.73±0.03	0.69±0.02	0.75±0.03
nine	4728	0.75±0.02	0.67±0.01	0.77±0.02
no	4750	0.69±0.04	0.67±0.02	0.73±0.03
off	4714	0.86±0.02	0.79±0.02	0.87±0.02
one	4740	0.75±0.04	0.63±0.02	0.76±0.04
on	4734	0.82±0.03	0.67±0.03	0.83±0.03
right	4734	0.68±0.02	0.62±0.03	0.71±0.02
seven	4754	0.76±0.02	0.64±0.02	0.79±0.02
sheila	3468	0.81±0.03	0.67±0.01	0.82±0.03
six	4738	0.86±0.03	0.85±0.02	0.89±0.03
stop	4760	0.80±0.02	0.73±0.02	0.81±0.03
three	4712	0.85±0.02	0.78±0.01	0.87±0.02
tree	3466	0.88±0.03	0.79±0.01	0.88±0.03
two	4746	0.86±0.02	0.65±0.02	0.86±0.02
up	4750	0.78±0.03	0.70±0.01	0.78±0.03
wow	3490	0.77±0.02	0.78±0.02	0.82±0.02
yes	4754	0.79±0.02	0.80±0.02	0.83±0.02
zero	4752	0.74±0.03	0.68±0.03	0.79±0.03
英単語ごとの平均値		0.79±0.01	0.71±0.01	0.81±0.01

各英単語の正解率を平均すると MFCC のみの特徴量では 0.79 ± 0.01 、デルタケプストラムのみでは 0.71 ± 0.01 、MFCC とデルタケプストラムの組み合わせでは 0.81 ± 0.01 となり、MFCC とデルタケプストラムの組み合わせで入力ベクトルを構成すると最も正解率が高くなった。ただし各英単語における平均値の評価は、英単語の 30 種類を標本としたときの 68% 信頼限界の標準誤差を用いた。デルタケプストラムのみを特徴量とした場合、MFCC のみの場合に比べて正解率は約 8% 低い。このことからサンプリングレートが 16 kHz における 1 秒間すべてを窓幅とした MFCC でも十分に正解率は望めるが、デルタケプストラムを加えた方がより正解率が向上することが分かる。

3.4.2 データセット数を説明変数とした性能評価値の変化

火山噴火波形のデータセット数は Speech Commands Dataset に比べてはるかに少ない。火山噴火波形のデータセット数を多くするためには、東郡元空振計における空振値が 10 Pa 以下のデータも集めてくる必要があるが、現在この自動化は実現していないために手で集めるコストが生じている。このことから火山噴火波形のデータセットの目標数を定める必要がある。ここでは、データセット数が性能評価値にどのように依存するかを調べる。手法としては Speech Commands Dataset のデータセット数を減らした場合に SVM で分類されたときの性能評価値の変動を見る。データセット数に対する目的変数は性能評価値の交差検証法で得られた平均値および標準誤差 (Standard Error:SE) の 2 種類である。また再現率と適合率の調和平均で F-値をとっているため、正解率と F-値のみで評価する。再現率と適合率におけるデータセット数の評価は付録に示す。ここで Speech Commands Dataset では英単語が 30 個あるが、任意のデータセット数に対する性能評価値の 1 点ごとに平均化し、平均値と標準誤差それぞれをグラフ化する。まずデータセット数に対する各英単語で得られた正解率の平均値を図 3.17 に示す。ただし、横軸は対数でプロットした。また F-値 (f1) についても同様にプロットしたグラフを図 3.18 に示す。

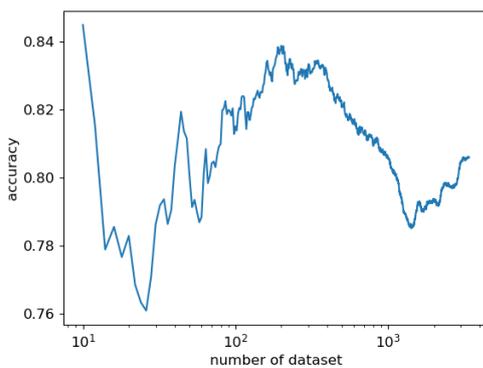


図 3.17 図：データセット数に対する各英単語の正解率

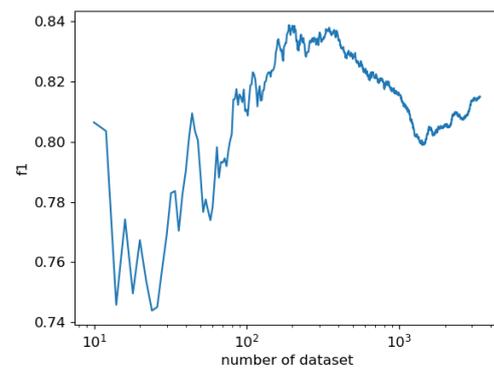


図 3.18 図：データセット数に対する各英単語の F-値

図 3.17 および図 3.18 について、データセット数に対する正解率と F-値はともにデータセット数が 10 から 20 にかけて急激に減少し、その後は上昇して約 200 でピークとなる。そこから約 1000 まで緩やかに減少し、その後は正解率、F-値ともに約 79% からまた上昇している。このことから各性能評価値にピークが存在していることが分かるため 4.1 節でピークについての考察をする。次にデータセット数に対する交差検証法によって得られた性能評価値の標準誤差の変化を示す。ただし、横軸は対数でプロットした。

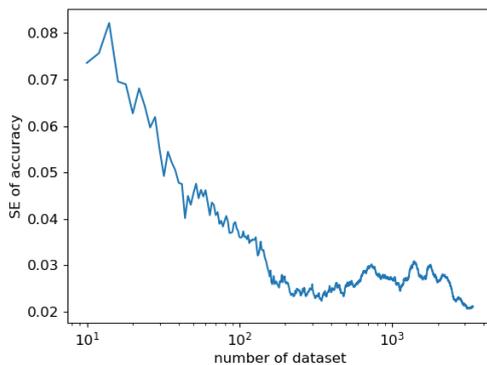


図 3.19 図：データセット数に対する交差検証法によって得られた正解率の標準誤差

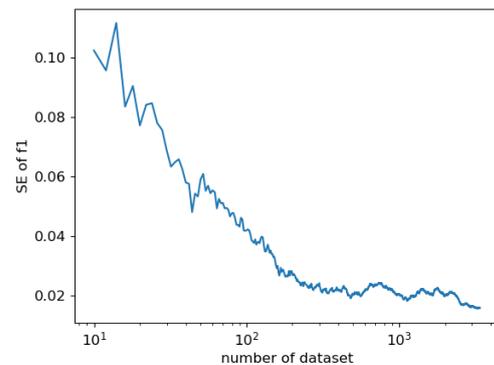


図 3.20 図：データセット数に対する交差検証法によって得られた F-値の標準誤差

データセット数 200 以上のときは正解率の標準誤差が 2% から 3% 程度しか変動しないのに対して、データセット数を 200 程度から少なくしていくと標準誤差が指数関数的に上昇し、データセット数が 20 以下になると正解率の標準誤差は 8% に到達した。このことから統計的に十分なデータセット数があるときは標準誤差は一定幅を保っているが、データセット数を減らしていくと標準誤差は急激に大きくなる。この急激に大きくなる境界を 4.1 節で考察する。

図 3.17 および図 3.18 から必ず性能評価値のピークが見つかる。このことから正解率と F-値を総合したときの最大性能評価値を求める。正解率を A 、F-値を F としたとき、本研究ではデータセット数 N のときのスコア平均 ($S_{ave} = (A + F)/2$) を用いて評価する。データセット数に対するスコア平均 S_{ave} の最大値と最大値が得られたときのデータセット数 N を表 3.4 に示す。

表 3.4 データセット数を説明変数としたスコア平均の最大値におけるデータセット数

英単語	データセット数	スコア平均の最大値
tree	150	0.99
six	136	0.99
off	234	0.98
bird	62	0.97
eight	734	0.96
two	308	0.96
yes	98	0.96
up	44	0.96
three	98	0.95
zero	58	0.95
on	38	0.95
stop	52	0.94
one	46	0.94
sheila	120	0.94
seven	22	0.93
four	48	0.91
go	12	0.90
dog	578	0.90
cat	980	0.88
bed	10	0.88
house	602	0.88
marvin	578	0.87
five	36	0.86
wow	90	0.86
down	10	0.86
right	10	0.86
left	12	0.86
no	10	0.85
happy	120	0.85
nine	304	0.82

表 3.4 ではスコア平均 S_{ave} の最大値について降順に並べている。各英単語の中で最もスコア平均 S_{ave} が高いのは tree と six であり、データセット数は tree が 150、six が 136 となった。この 2 単語における性能評価値とデータセット数を火山噴火波形データセットと比較し、4.3 節で考察する。

3.5 火山噴火波形の特徴量抽出および分類

3.5.1 相関値の閾値倍率を説明変数とした性能評価値の変化

データセット数候補数 N_s は 3.3 節によれば閾値倍率 t によって定まる。ただし、訓練データとテストデータを 5 分割して 4:1 で割り当てているため、テストデータセット数が 0 になる場合は性能評価値の計算ができなくなる。そのため、閾値倍率 t の最大値は各観測点によって異なることに注意する。今回、性能評価は正解率 (accuracy)、再現率 (recall)、適合率 (precision)、そして再現率と適合率の関係性を評価する F-値 (f1) の 4 種類の値を用いて性能評価値を計測することにした。閾値倍率 t に対する各性能評価値を図 3.21 から図 3.36 にそれぞれ示す。ただし特徴量は MFCC とデルタケプストラムの 2 種類の組み合わせで構成し、計 24 次元である。

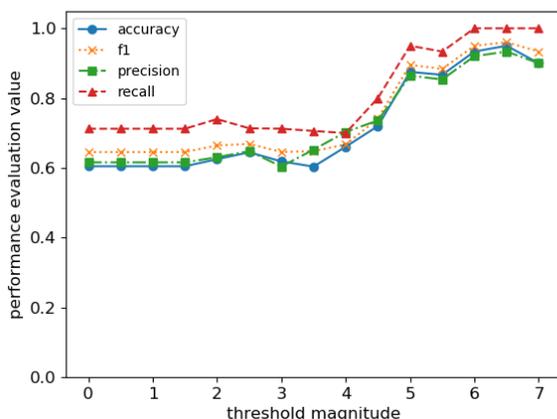


図 3.21 香美市土佐山田町宮ノ口における相関値の閾値倍率に対する性能評価値

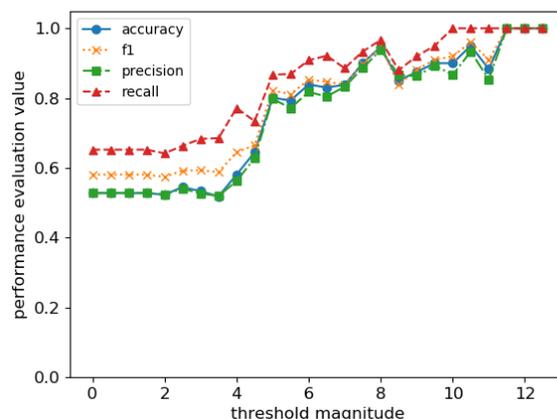


図 3.22 黒潮町浮鞭における相関値の閾値倍率に対する性能評価値

図 3.21 から図 3.36 において、横軸はテストセット数が 0 以外の相関値の閾値倍率 t であり、縦軸は各性能評価値である。各性能評価値はすべて 0 から 1 までの値をとる。3.3 節と比較したとき、閾値倍率が約 3 まではデータセット候補数 N_s が一定であるために、性能評価値もそれぞれ一定となってる。さらに高い閾値倍率 t の場合には、土佐清水市足摺岬 1 と土佐清水市足摺岬 2 以外では、どの観測地においても性能評価値が向上している。各観測点における閾値倍率 t の限界値付近はテストデータが非常に少ない場合に相当するが、観測地によっては各性能評価値は上下する傾向にある。

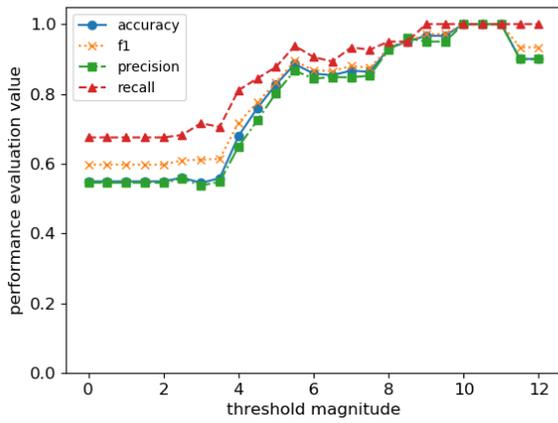


図 3.23 黒潮町蜷川における相関値の閾値倍率に対する性能評価値

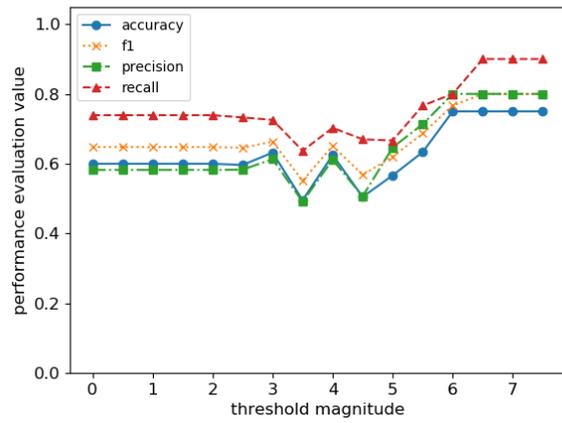


図 3.24 黒潮町上川口における相関値の閾値倍率に対する性能評価値

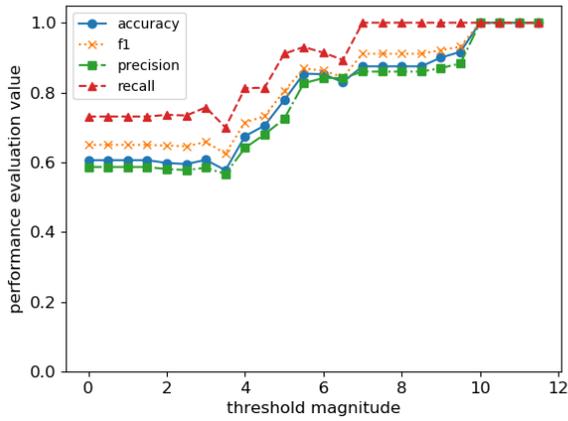


図 3.25 黒潮町出口における相関値の閾値倍率に対する性能評価値

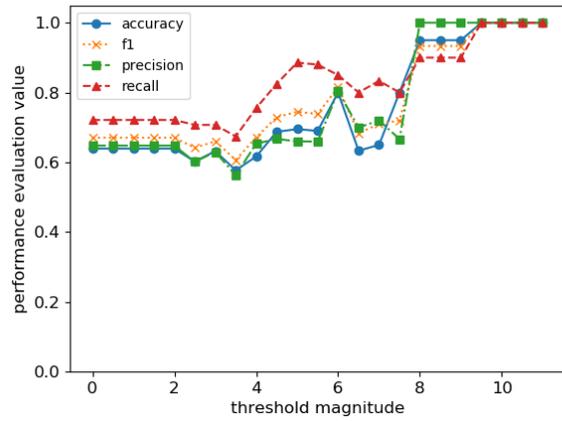


図 3.26 黒潮町馬荷における相関値の閾値倍率に対するデータセット候補数の変化

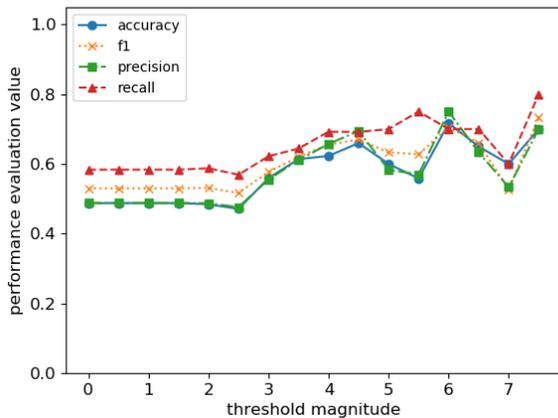


図 3.27 宿毛市小筑紫町田ノ浦における相関値の閾値倍率に対する性能評価値

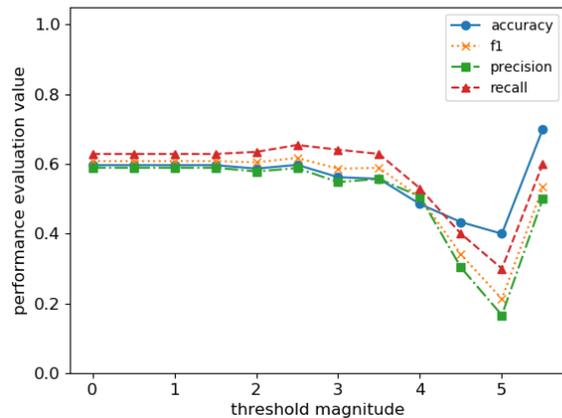


図 3.28 土佐清海市足摺岬 1 における相関値の閾値倍率に対する性能評価値

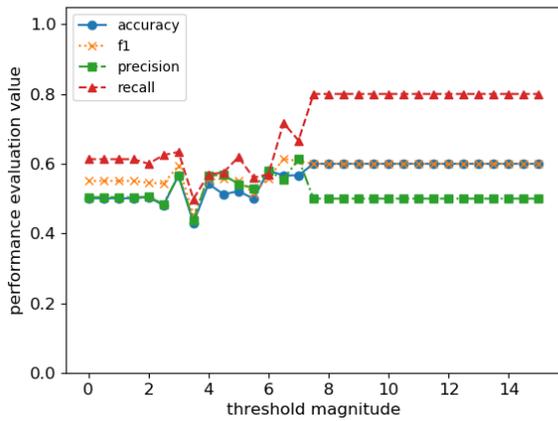


図 3.29 土佐清水市足摺岬 2 における相関値の閾値倍率に対する性能評価値

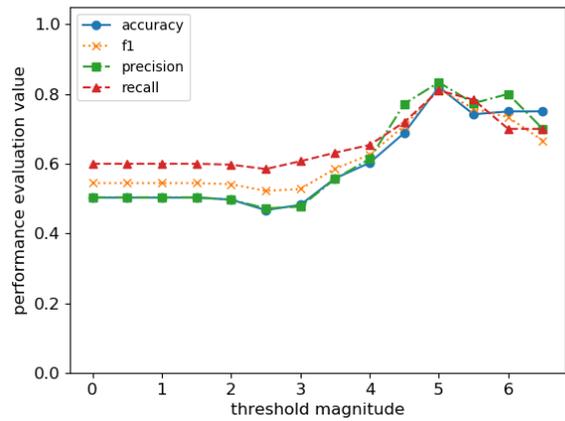


図 3.30 土佐清水市足摺岬 3 における相関値の閾値倍率に対する性能評価値

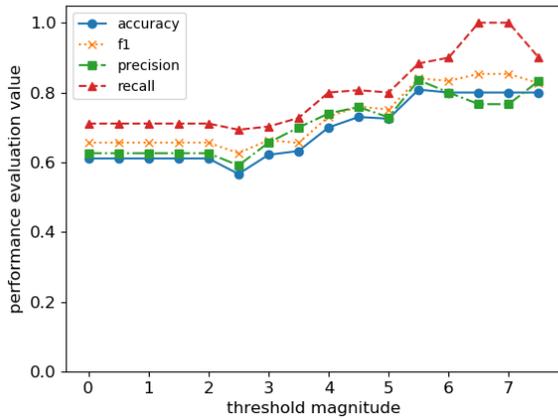


図 3.31 土佐市宇佐町宇佐における相関値の閾値倍率に対する性能評価値

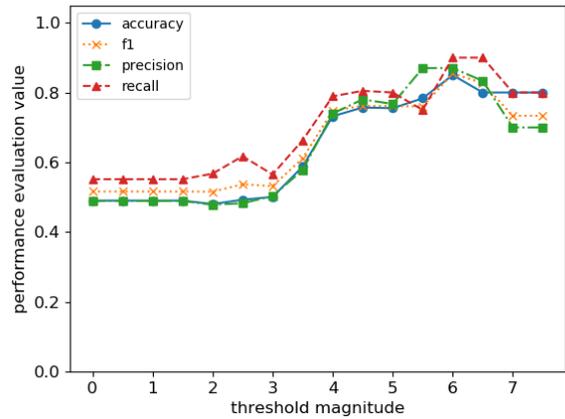


図 3.32 高知市春野町芳原における相関値の閾値倍率に対する性能評価値

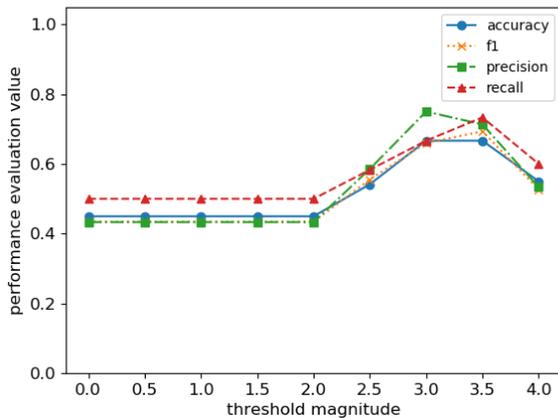


図 3.33 南国市物部における相関値の閾値倍率に対する性能評価値

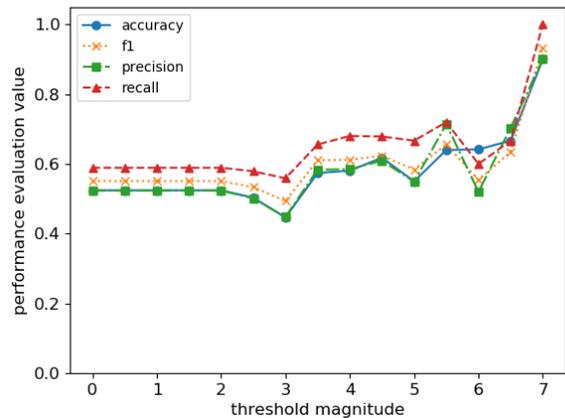


図 3.34 安芸市西浜における相関値の閾値倍率に対する性能評価値

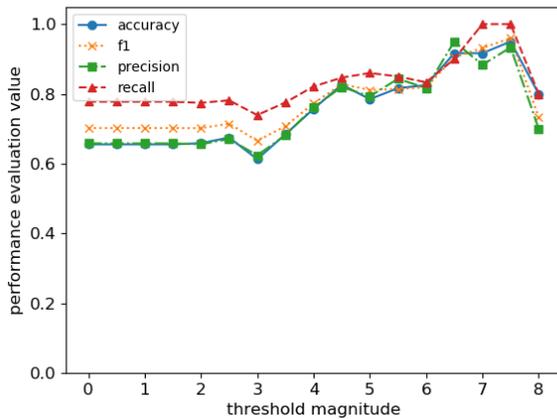


図 3.35 東洋町生見における相関値の閾値倍率に対する性能評価値

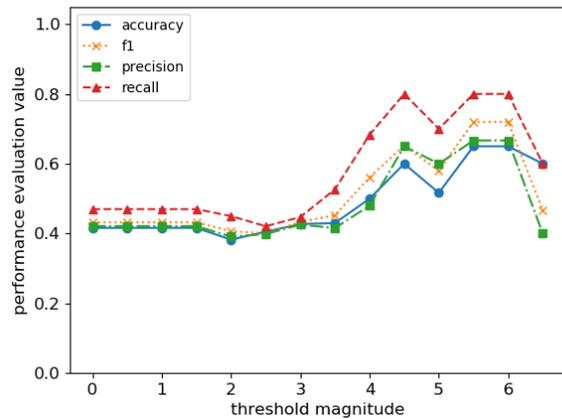


図 3.36 室戸市室戸岬町における相関値の閾値倍率に対する性能評価値

3.5.2 MFCC またはデルタケプストラムを特徴量としたときの最大性能評価値

3.5.1 項では特徴量を MFCC とデルタケプストラムの 2 種類で構成したが、火山噴火波形でもデルタケプストラムを用いて構成した特徴量を入力してよいかを判断するために、MFCC のみ、デルタケプストラムのみ、MFCC とデルタケプストラムで特徴量を構成した場合を比較する。ここで、正解率を A 、F-値を F としたとき、これらを平均したスコア平均 $S_{ave} = (A + F)/2$ で比較する指標とする。各性能評価値の最大値は 1 であるが、各性能評価値が 1 の場合は恣意的なテストセットとなっていると考えられる。そのため、正解率と F-値の平均するときに、各性能評価値のいずれかが 1 の場合を除くという条件を加える。前述の条件のときの各閾値倍率 t におけるスコア平均 S_{ave} の最大値とそのときの閾値倍率 t の比較を表 3.5.2 に示す。

特徴量が MFCC とデルタケプストラム (MFCC+ Δ) の組み合わせで構成された場合で、スコア平均 S_{ave} が 90% を超えているのは黒潮町蜷川、黒潮町浮鞭、東洋町生見であり、黒潮町蜷川と黒潮町浮鞭はそれぞれ 95% ($t=8.5$, $t=8$) となっている。最も低いのは土佐清水市足摺岬 2 であり 60% である。また MFCC のみに比べて MFCC+ Δ の方が黒潮町蜷川において 2% だけスコア平均 S_{ave} が上昇しており、下降している観測点は見られない。デルタケプストラムのみを特徴量とした場合は、MFCC のみに比べて土佐清水市足摺岬 2 以外ではいずれも性能評価値は減少している。これから英単語と同様に MFCC+ Δ の組み合わせが最も性能評価値が高くなり、MFCC のみでも十分な性能評価値となる。4.3 節では最もスコア平均 S_{ave} が高くなった黒潮町蜷川と黒潮町浮鞭について、同様にスコア平均 S_{ave} が高くなった英単語 tree と six とを比較する。

表 3.5 MFCC またはデルタケプストラム (Δ) における閾値倍率 t のスコア平均 S_{ave} の最大値とその閾値倍率 t

観測地	スコア平均 S_{ave} の最大値			最大時の閾値倍率 t		
	MFCC	Δ	MFCC+ Δ	MFCC	Δ	MFCC+ Δ
安芸市西浜	0.65	0.64	0.65	6.5	7	6.5
香美市土佐山田町	0.89	0.81	0.89	5	6	5
高知市春野町	0.85	0.66	0.85	6	7	6
黒潮町出口	0.86	0.74	0.86	6	10	5.5
黒潮町上川口	0.77	0.55	0.77	6.5	0	6.5
黒潮町蜷川	0.93	0.77	0.95	8	11.5	8.5
黒潮町浮鞭	0.95	0.86	0.95	8	11.5	8
黒潮町馬荷	0.81	0.62	0.81	6	5.5	6
室戸市室戸岬町	0.69	0.52	0.69	5.5	5	5.5
南国市物部	0.68	0.55	0.68	3.5	3	3.5
宿毛市小筑紫町	0.72	0.60	0.72	7.5	5.5	7.5
土佐清水市足摺岬 1	0.62	0.54	0.62	5.5	2.5	5.5
土佐清水市足摺岬 2	0.60	0.61	0.60	3	7.5	7.5
土佐清水市足摺岬 3	0.82	0.66	0.82	5	5	5
土佐市宇佐町宇佐	0.82	0.80	0.82	5.5	7	5.5
東洋町生見	0.91	0.85	0.91	6.5	7.5	6.5

4 考察・評価

4.1 データセット数の最適値

各英単語のデータセットについて、データセット数 N を説明変数としたときの各英単語の交差検証法を用いて線形 SVM にかけたときの正解率と F-値の平均値および標準誤差の変化をグラフ (図 3.17 から図 3.20) にすることができた。このグラフから最適なデータセット数 N_{opt} を評価し、火山噴火波形のデータセット数 N がどれだけ必要なの考察する。正解率と F-値がピークとなったときのデータセット数を表 4.1 に示す。ただし性能評価値が計算可能な下限データセット数付近であるときにはこれらの性能評価値は著しく不安定になる。そのため、 $N \geq 20$ のときのみを評価する。

各性能評価値のピーク時のデータセット数 N_{max} について、正解率と F-値はともに 190 となった。これを用いれば各英単語の最適なデータセット数は $N_{max} = 190$ と定まる。一

表 4.1 各英単語のデータセット数に対する性能評価値が最大となるときのデータセット数 N_{max}

性能指標名	性能評価値	データセット数 N_{max}
正解率	0.84	190
F-値	0.84	190

方、交差検証法を用いた場合の標準誤差について、データセット数 N を大きくすると、急激に減少し一定値となったと見做せる。この一定値はデータセット数 $N = 3426$ に対する標準誤差の集合の統計量である平均値 μ 、標準偏差 σ に対して、 $\mu \pm \sigma$ の幅であると定義する。一定幅の大きい方 $\mu + \sigma$ を閾値として、データセット数を減少させたときその閾値を超えたときを少なくとも必要なデータセット数 N_{need} とする。この閾値のビジュアル化を図 4.1 および図 4.2 に示す。また性能評価値それぞれにおける必要データセット数 N_{need} について表 4.2 に示す。

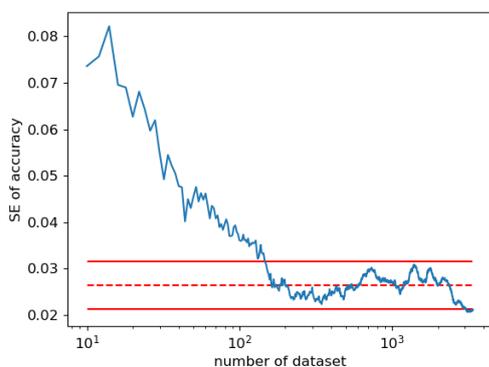


図 4.1 データセット数に対する正解率の標準誤差において、閾値による必要データセット数の箇所

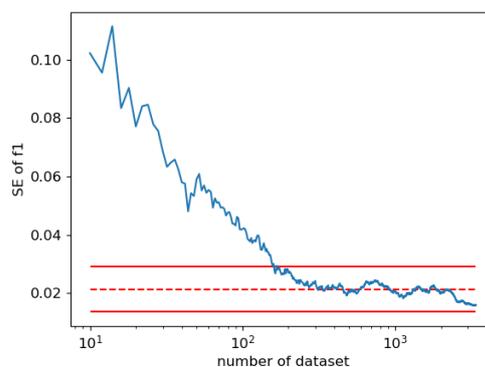


図 4.2 データセット数に対する F-値の標準誤差における閾値による最小データセット数の箇所

表 4.2 データセット数に対する性能評価値において交差検証法による標準誤差が閾値を超えたときの必要データセット数 N_{need}

性能指標名	標準誤差	データセット数 N_{need}
正解率	0.03	146
F-値	0.03	168

各性能評価値のピーク時の議論と同様に、交差検証法による標準偏差における必要データセット数 N_{need} を求める。表 4.2 より 146 から 168 のところに標準誤差が安定するための必要データセット数 N_{need} があると考えられる。本研究では交差検証法による標準誤差に

おける各性能評価値で得られた必要データセット数の最小値を $N_{need} = 146$ とする。性能評価値のピークおよび標準誤差の一定値を超えたときの二要素をそれぞれ考えた。総じて考えたとき、性能評価値の向上のために最適なデータセット数 N_{opt} は $146 \leq N_{opt} \leq 190$ の範囲にあると言えるため、両者を平均した $N_{opt} = 168$ を最適なデータセット数と定める。ここで、訓練データが多ければ多いほど分類の正解率が上がることを仮説とした場合、疑問点としてはデータセットを増加させても実際には性能評価値の向上にはつながっていないことにある。統計学的に当然のことかもしれないが、今回これは全体データに対して訓練データの割合が増えすぎており、良好な線形識別関数の学習には至らないことを示していると考えられる。しかし、交差検証法の標準誤差については説明可能である。データセット数が多くなると SVM の線形識別関数を学習により決定する際、データセットに含まれていた異例なデータに依存されなくなるので、データセット数が多くなると標準誤差は小さくなる。データセット数に対する性能評価値のピークが現れることの説明がつかないことからデータセット数の最適値 N_{opt} ではなく必要データセット $N_{min} = 146$ としてデータセット数を評価してもよい考えられる。

次にインフラサウンド領域における火山噴火波形について考える。性能評価値の向上および標準誤差の安定を目的とするのであれば、上述より SVM を分類器とした場合にデータセット数は $N_{opt} = 168$ 付近であればよい。火山噴火波形のデータセット候補数 N_s は閾値倍率 t に依存する。今回は閾値倍率 t を説明変数とした場合における火山噴火波形かノイズ波形の分類をしたときのスコア平均 S_{ave} が最大値となったときのデータセット数 N_{infs} を表 4.3 に示す。

表 4.3 で明らかな通り、現状ではすべての観測点について、 $N_{infs} \leq N_{opt}$ である。このことから、観測期間や観測点を増やすことでさらにデータセット数を増やす必要がある。本研究で用いた期間内 (2016 年 8 月 13 日から 2018 年 12 月 31 日) であれば火山噴火波形のデータセット数を増やすには気象庁の東郡元空振計において 10 Pa 未満の弱い火山噴火も参照し、対応するインフラサウンドセンサーの観測データを取得するしかない。本研究でデータサーバーに保管されている観測データをウェブサイトからダウンロードする形式で収集したため手動といえる。自動収集には火山噴火のタイムスタンプを用いて FPT サーバーからダウンロードするプログラムなどを構築する必要がある。このシステムは平田 (2019) により既に実装済み [34] であり、これを改良することにより任意の時刻における火山噴火波形を取得できる。

4.2 閾値変化によるデータセット破棄数の理論化

相関値の閾値倍率 t の変化によるデータセット候補数 $N_s(t)$ の結果は 3.3 節で述べた。特徴的なのは、閾値倍率 t が約 3 を超えるとデータセット候補数 $N_s(t)$ が急激に落ちるが、閾値倍率 t が約 5 付近で緩やかになることであった。データセット破棄数 N_s が急激に落ちて緩やかになる傾向をロジスティック曲線でフィッティングさせることを提案する。ロジス

表 4.3 火山噴火波形におけるスコア平均が最大値のときのデータセット数

観測地	データセット数 N_{infs}
安芸市西浜	30
香美市土佐山田町	36
高知市春野町	24
黒潮町出口	84
黒潮町上川口	12
黒潮町蜷川	38
黒潮町浮鞭	58
黒潮町馬荷	42
室戸市室戸岬町	16
南国市物部	22
宿毛市小筑紫町	10
土佐清水市足摺岬 1	10
土佐清水市足摺岬 2	16
土佐清水市足摺岬 3	44
土佐市宇佐町	34
東洋町生見	26

ティック曲線は数理生物学の分野に現れる。現実世界の個体群動態の個体数は指数関数的に増加とされるマルサスモデルがあるが、マルサスモデルに個体数の過剰増加を抑制する自己制御過程を導入したロジスティック型増殖モデルがロジスティック曲線となる [35]。本研究で得られた閾値倍率 t のデータセット候補数 $N_s(t)$ に対して、ロジスティック曲線で見られる S 字形状とは変曲点を中心に対称である。ここで閾値倍率 t を説明変数としたデータセット候補数 $N_s(t)$ について、初期データセット候補 $N_s(0)$ から相互相関関数によってどれだけのデータが破棄されたかを評価するために、データセット破棄数 $N_d(t) \equiv N_s(0) - N_s(t)$ として考える。ただし以下では簡易のために $N_d = N_d(t)$ と略記する。ここで次の二つの仮定を導入する。1.) データセット破棄数の変化率 $\frac{dN_d}{dt}$ はデータセット破棄数 N_d に比例する。すなわち、定数を m とするとき

$$\frac{dN_d}{dt} = mN_d \quad (12)$$

と表される。この仮定は数理生物学におけるマルサスモデルと同様であり、本研究ではデータセット破棄数 N_d は閾値倍率 t が増加するにしたがって指数関数的に N_d が増加することを意味する。2.) データセット破棄数 N_d が増加すると、増加率が抑えられる。このときの増

加率を $m(N_d)$ とおく。比例定数を r 、定数を K としたとき、 $m = -\frac{r}{K}N_d + r = \frac{r}{K}(K - N_d)$ と仮定する (図 4.3)。

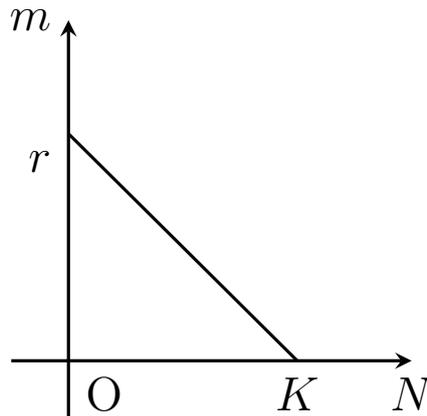


図 4.3 データセット破棄数 N_d とデータセット破棄増加率 m の関係

このとき、(12) 式に 2. で仮定した式を代入すると次の微分方程式が得られる。

$$\frac{dN_d}{dt} = \frac{r}{K}(K - N)N \quad (13)$$

(13) 式はロジスティック方程式と呼ばれる。(13) 式から次の解が解析的に求められる。

$$N = \frac{KN_0}{N_0 + (K - N_0)e^{-rt}} \quad (14)$$

ただし、 N_0 は閾値倍率 $t = 0$ における初期データセット破棄数付近のパラメータであり $N_0 \neq 0$ であるが、今回の場合は $N_0 \approx 0$ をとる。定数 K は閾値倍率 t が十分大きいときにデータセット破棄数が収束する値となる。今回は閾値倍率 t が約 5 付近で緩やかになったときの N_d を K とする。また (13) 式をさらに 2 階微分した 3 階微分方程式

$$\frac{d^3N}{dt^3} = -\frac{2r}{K} \quad (15)$$

の左辺は数値的に求められるので、比例定数 r が推定できる。しかし、オイラー法で解く場合には刻み幅を小さくする必要があり [36]、刻み幅を大きくすると誤差が大きくなることが知られている。そのため本研究において r の推定が困難であるので、 r もパラメータと同様の扱いをする。このとき黒潮町浮鞭データセットにおいて今回の理論値と実験値を比較したグラフを図 4.4 に示す。ここで $K = 349$ ($t = 6$)、 $r = 2$ 、 $N_0 = 0.2$ とした。 $t = 6$ までは実験値とおよそ一致する。しかし、収束する値を $t = 6$ で指定しているので、それ以降は一致しない。ここで $t = 6$ 以降を拡大してプロットしたグラフを図 4.5 に示す。

直線的に見えるが、閾値倍率が $t = 7.5, 10, 12$ を境に傾きが緩やかになっている。仮説として、ロジスティック曲線による始めのデータセット破棄数の増加率は大きいですが、その後、何度も繰り返して、破棄数限界 $N_s(t) = 0$ に到達するのではないかと考えられる。繰り返

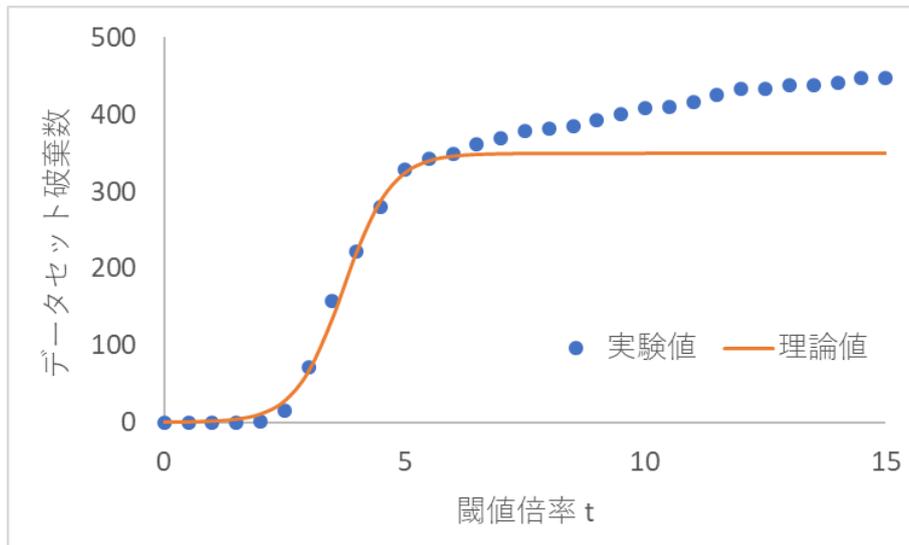


図 4.4 黒潮町浮鞭における閾値倍率 t に対するデータセット破棄数 N_d の理論値との比較

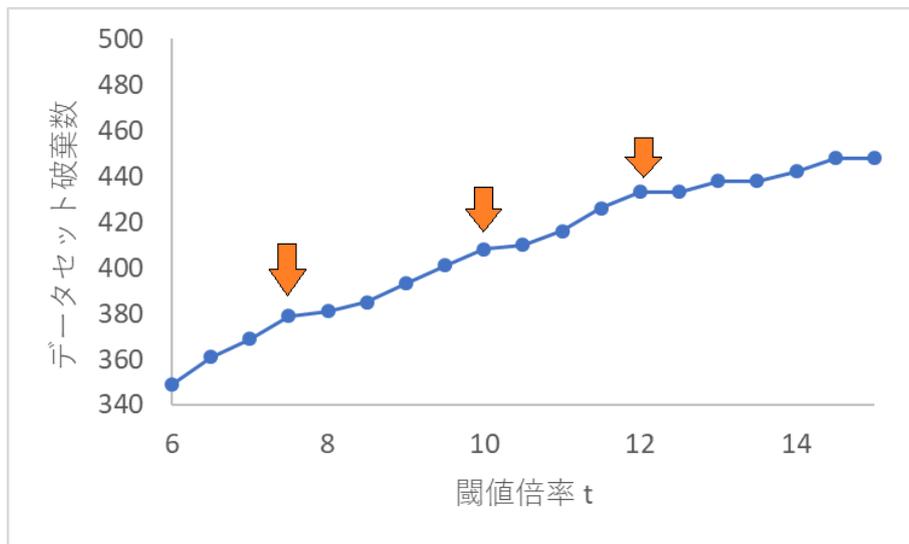


図 4.5 黒潮町浮鞭における閾値倍率 $t = 6$ 以降のデータセット破棄数。矢印は破棄数増加率が緩やかになっている部分

返し破棄する曲線が描かれると考えると、火山噴火波形候補として選んできた波形データは繰り返した曲線の数と同じ種類が選別できるのではないかと考えられる。この結果から総括的に言えることは、繰り返し発生するような大規模自然災害について、音波リモートセンシングによって取得した波形データ群にはデータの欠損および高いノイズレベルによって破棄されなければならないデータが含まれおり、相互相関処理によって破棄した場合、その破棄数はロジスティック曲線に従うと考えてよい。

表 4.4 最もスコア平均 S_{ave} が高い英単語と火山噴火波形データセットの性能評価値およびデータセット数の比較

	データセット数	正解率	再現率	適合率	F-値
tree	150	0.987±0.007	0.99±0.01	0.99±0.01	0.987±0.007
six	136	0.985±0.008	0.99±0.01	0.99±0.01	0.985±0.008
黒潮町蜷川	38	0.950±0.027	0.95±0.04	0.96±0.03	0.950±0.027
黒潮町浮鞭	58	0.946±0.019	0.96±0.03	0.93±0.03	0.946±0.019

4.3 可聴音と火山噴火波形との比較

可聴音の音声である Speech Commands Dataset の英単語とインフラサウンドセンサーから取得した火山噴火波形を比較する。両者とも MFCC およびデルタケプストラムで特徴量抽出を行い、スコア平均 S_{ave} が最も高かったデータセットを 3.4.2 項、3.5.2 項それぞれで示した。英単語では tree と six のデータセットが最も高くなり、火山噴火波形データセットでは黒潮町蜷川と黒潮町浮鞭が最も高くなった。それぞれのデータセットで得られたときの正解率、再現率、適合率、F-値およびそのときのデータセット数を表 4.3 に示す。

tree と six はスコア平均 S_{ave} は同じであるがデータセット数が多く、標準誤差が小さい tree を火山噴火波形データセットとの比較対象は tree とする。黒潮町蜷川と黒潮町浮鞭についても同様に、データセット数が多く、標準誤差が小さいことから黒潮町浮鞭を英単語 tree との比較対象とする。英単語 tree のデータセットと黒潮町浮鞭における火山噴火波形データセットを比較すると、特に問題になるのが黒潮町浮鞭におけるデータセット数の少なさと、標準誤差が高くなっていることである。4.1 節で述べたように標準誤差の高さはデータセットを増やせば解決する。そのことから黒潮町浮鞭における火山噴火波形のデータセットをさらに増やことを目指す必要がある。総じてデータセット数の少なさは課題となるが閾値倍率 $t = 8.0$ における黒潮町浮鞭データセットが火山噴火データセットの中で最も性能評価値が高く、良質なデータセットであると考えられる。最後に黒潮町浮鞭データセットがどのようなデータセットの構成になっているかを述べる。黒潮町浮鞭データセットにおける生データを前処理した波形データについて、真クラス p^*, n^* の個数 $P = N = 29$ のうち、 $P = N = 10$ だけそれぞれ示す。ここで火山噴火波形 p^* にある各色の垂直線については 2.3.2 項で述べたのと同様である。

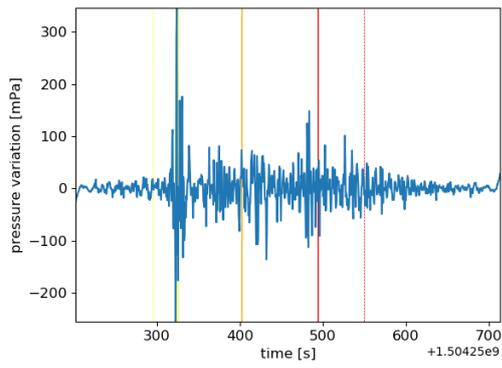


図 4.6 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 1

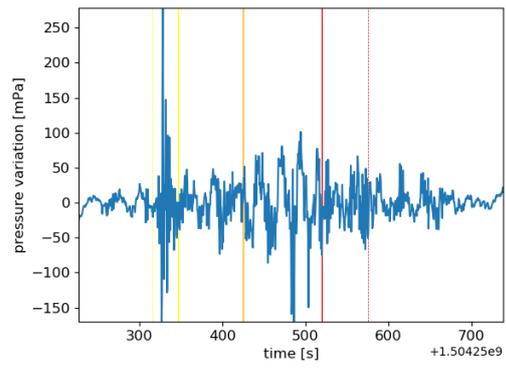


図 4.7 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 2

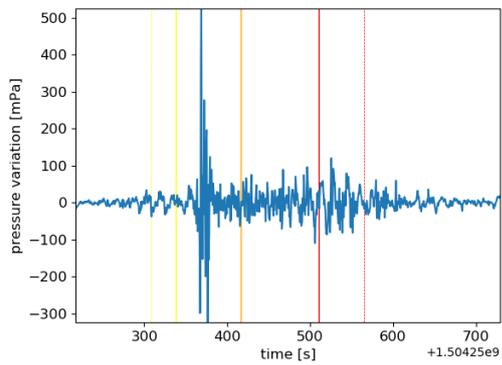


図 4.8 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 3

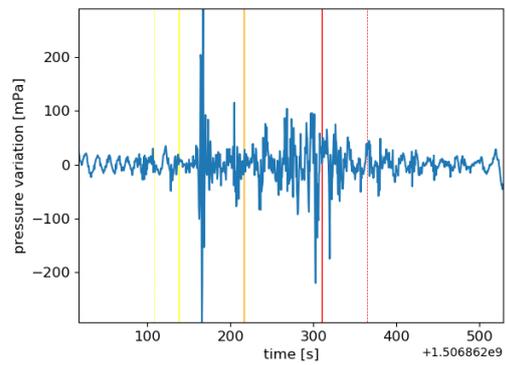


図 4.9 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 4

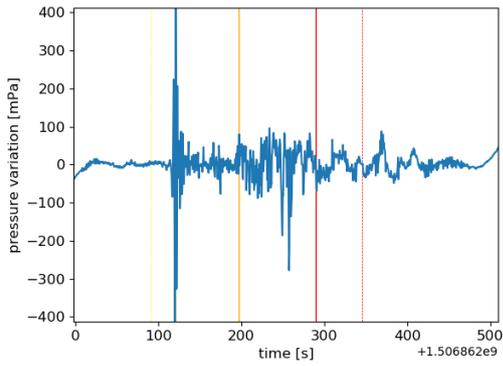


図 4.10 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 5

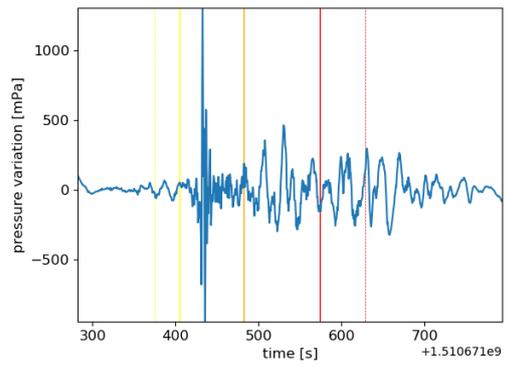


図 4.11 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 6

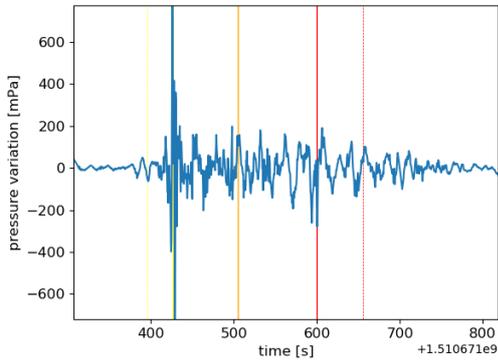


図 4.12 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 7

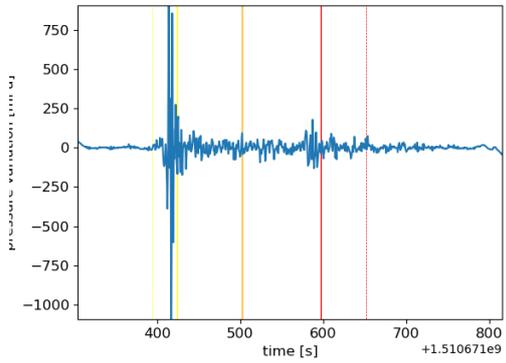


図 4.13 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 8

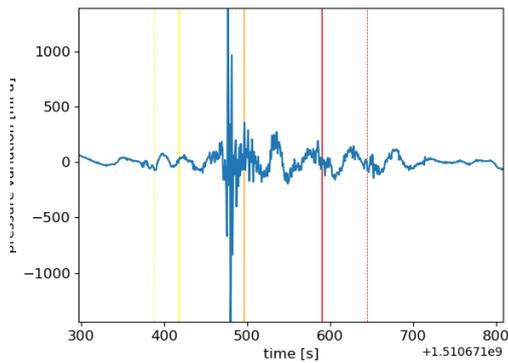


図 4.14 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 9

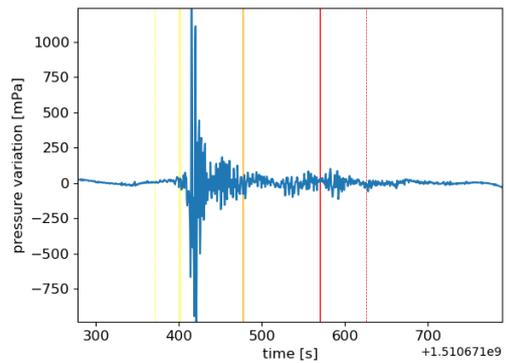


図 4.15 黒潮町浮鞭データセットにおける火山噴火波形 p^* の前処理された波形データ 10

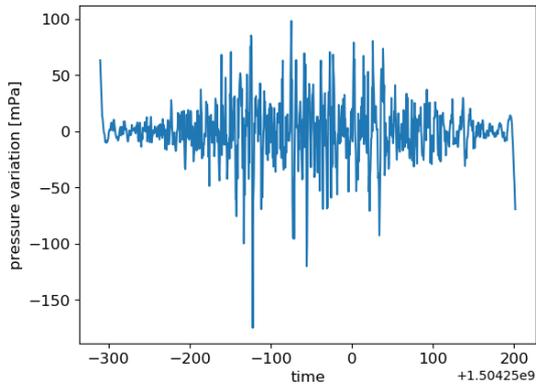


図 4.16 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 1

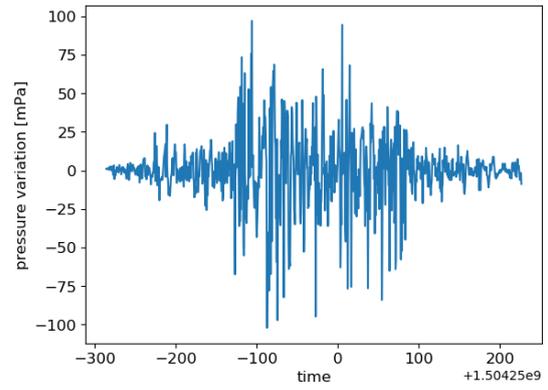


図 4.17 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 2

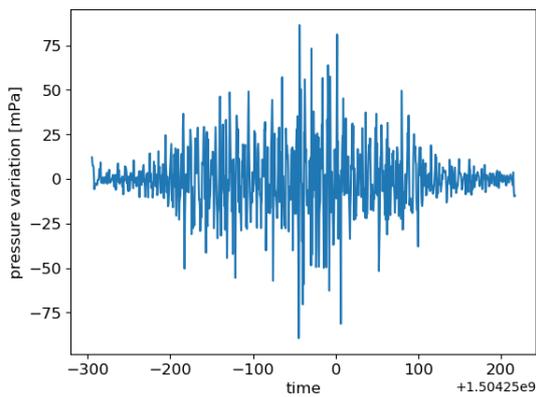


図 4.18 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 3

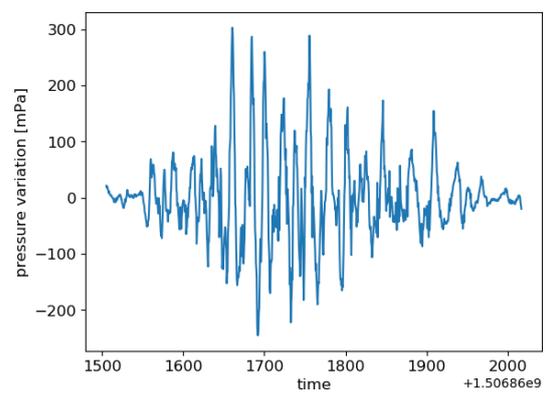


図 4.19 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 4

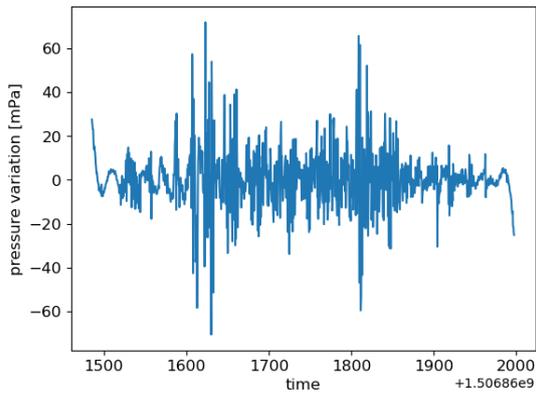


図 4.20 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 5

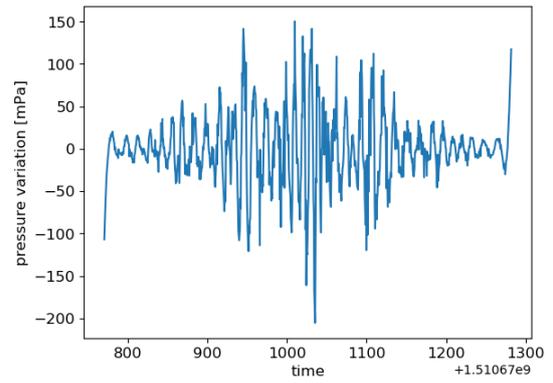


図 4.21 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 6

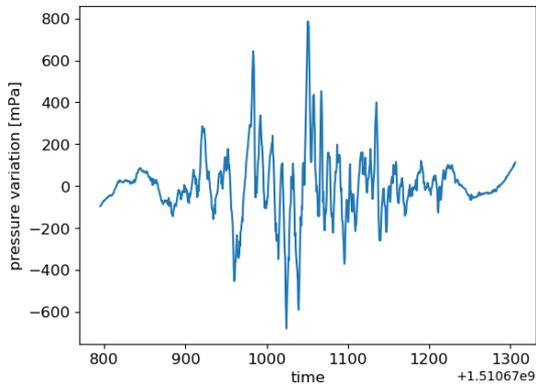


図 4.22 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 7

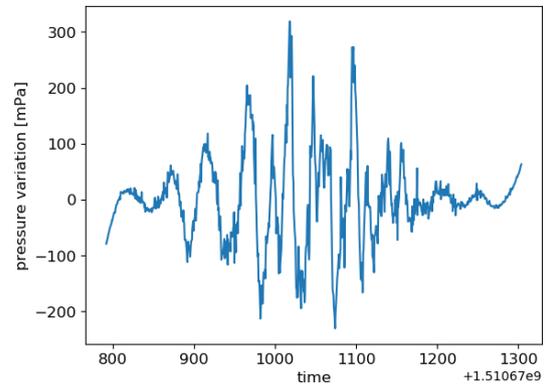


図 4.23 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 8

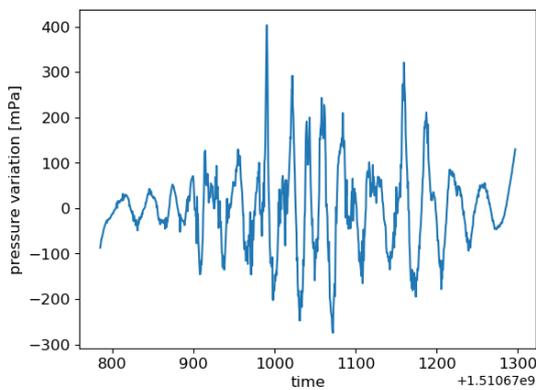


図 4.24 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 9

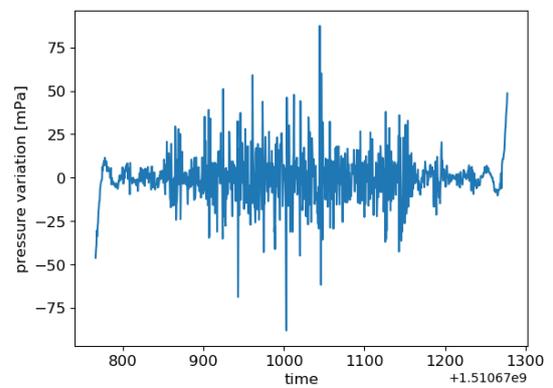


図 4.25 黒潮町浮鞭データセットにおけるノイズ波形 n^* の前処理された波形データ 10

火山噴火波形 p^* とノイズ波形 n^* を比較した場合、火山噴火波形が到達したことによる気圧変動のピークとなる規模は各火山噴火波形において低くても 250 mPa 程度ある。それに比べてノイズ波形 n^* ではハミング窓がかかっていることが確認でき、ノイズが時系列全体にちらばっていることがわかる。しかし、火山噴火波形 p^* ではピークが到達予測時刻内に収まっているため、ハミング窓を掛けている様子は波形から確認しにくいことから、 p^* 、 n^* が区別できる。これらは目視で p^* 、 n^* を区別できるが、中には区別できない波形データが含まれていた。本論文執筆時点においては SVM で区別されたとはいえないことに注意したい。

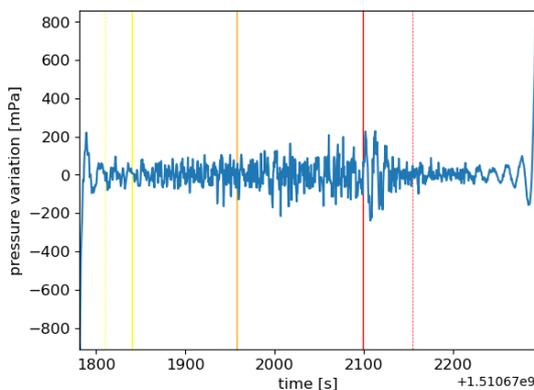


図 4.26 黒潮町浮鞭データセットにおける偽陰性と考えられる波形データ

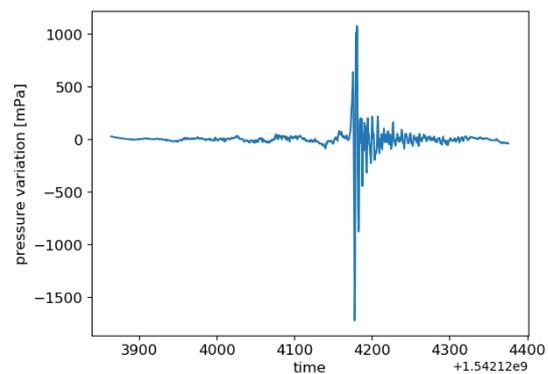


図 4.27 黒潮町浮鞭データセットにおける偽陽性と考えられる波形データ

図 4.26 について、両端が相対的に振幅が高いが、それを除けばノイズレベルが全体的に高く火山噴火が見られないような波形である。ゆえに偽陰性と判断されたと考えられる。次に図 4.27 であるが、ノイズレベルが低い状態から突発的に 1000 mPa ほどの強烈な空振しており、火山噴火波形と酷似している。この波形は 2018 年 11 月 14 日 0 時 43 分に発生した火山噴火のインフラサウンド波形が小筑紫町田ノ浦のセンサーに到達する手前の波形データであるために n^* とされた。 n^* となった理由には 2 つ考えられる。インフラサウンド波形到達予測よりも早く火山噴火が到達した、もしくはそもそも火山噴火波形ではなく雷で発生した波形であるということである。この場合の火山噴火は規模が比較的大きく、黒潮町蜷川でも 2000 mPa を達しており (図 4.28)、予想よりも早くインフラサウンド波形が到達したと考えられる。しかし、その後も到達予想時刻の範囲において 150 mPa 程度の空振が確認されており (図 4.29)、黒潮町浮鞭データセットに含まれているため、波源を特定することを結論付けるのは困難である。

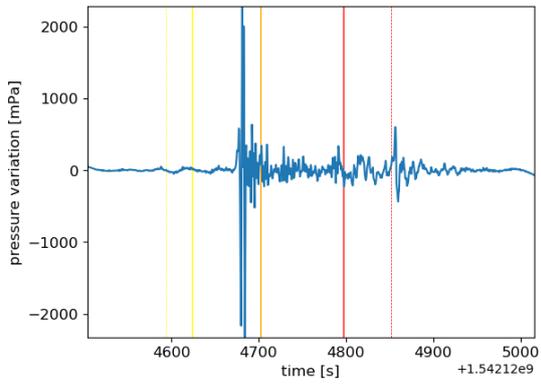


図 4.28 2018 年 11 月 14 日 0 時 43 分に火山噴火において黒潮町蜷川で取得したインフラサウンド波形

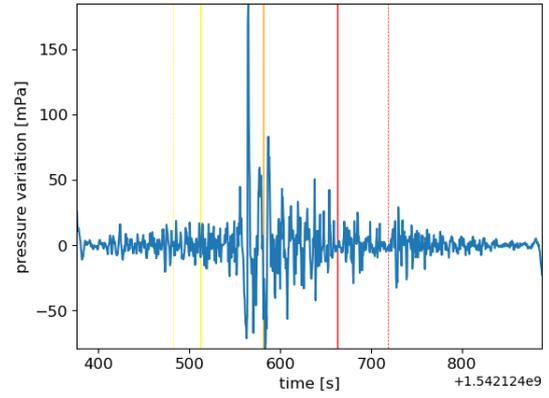


図 4.29 2018 年 11 月 14 日 0 時 43 分に火山噴火において小筑紫町田ノ浦で取得したインフラサウンド波形

5 結論

火山噴火が発生してからインフラサウンドをリモートセンシングした地球物理学的イベントが機械学習アルゴリズムによって分類できるか検証を行った。今回は訓練データを数多く集めることができることと、高知県内の観測地点までへの到達が予測できることから、桜島の火山噴火をターゲットにした。ここで観測地点への音波到達を判断するために、一回の火山噴火に対して各観測点の波形との相互相関関数を取り相関値の閾値を高める手法を提案した。火山噴火波形と火山噴火波形でないノイズ波形の2クラス分類のためのデータセットについて、MFCC とデルタケプストラムで 24 次元の特徴ベクトルを構成し線形 SVM に入力したとき、相関値の閾値倍率 t を高めるほど性能評価値が向上することが確認された。今回解析で用いた 16 か所のうち、スコア平均が最も高いインフラサウンド観測点は高知県黒潮町浮鞭であり、正解率、再現率、適合率、F-値はそれぞれ 0.95 ± 0.02 、 0.96 ± 0.03 、 0.93 ± 0.03 、 0.95 ± 0.02 となり、そのときのデータセット数は 58 であった。このデータセット数に対して Speech Commands Dataset の英単語 30 種類を統計的に十分なデータとして比較対象とした。各単語それぞれに 1600 を超える音声データ群が得られる Speech Commands Dataset データに対して性能評価値が最大となるデータセット数の最適値は 168 と推定でき、黒潮町浮鞭におけるこれまでに観測された火山噴火波形のデータセット数は少なく、学習される統計量に対していまだ十分でないと判断した。現在のところ気象庁が公開している火山噴火発生時刻に対する一部分しか収集できていないため、データセットを自動的に集め、観測の継続とともにデータセット数を蓄えていく手段を開発することが今後の課題となる。

謝辞

本研究のコンセプトであるインフラサウンド波形を機械学習の手法を用いて分類し、大規模自然災害に対する防災・減災に役立てるという課題の指導を引き受けてくださり、修士課程のを2年間および学士のアドバンスプログラムによる2年次仮配属から数えて計3年間の長期間に渡ってご指導ご鞭撻を賜りました高知工科大学 システム工学群 山本 真行教授には深く感謝申し上げます。また本論文の副査を引き受けてくださった星野 孝総 准教授、小林 弘和 准教授をはじめとするシステム工学群 電子系の先生方、には学士、修士の両課程にかけて講義などを通して数多くのことを学ばせて頂いたことを感謝するとともに深くお礼を申し上げます。本研究ではインフラサウンド波形の特徴量抽出を可聴音と比較して評価をするために Google 社が Creative Commons BY 4.0 ライセンスで提供している the Speech Commands Dataset を利用させて頂きました [37]。これにより音声認識の基礎的な手法を学ぶことができ、その手法をインフラサウンドデータ研究に応用させることができました。Speech Commands Dataset を開発した TensorFlow チームと AIY チームに深く感謝申し上げます。さらに AIY のウェブサイトを通して英単語を発声していただいた数千人の皆様にも感謝申し上げます。本研究では火山噴火のインフラサウンドとその大気中伝播過程に関する知識が必要なときにご指導を賜りましたシステム工学群 齊藤 大晶 助教 (現北海道大学 大学院理学研究院 博士研究員) に深く感謝申し上げます。また研究室データサーバーを構築した研究室 OB の平田 大祐 様、現在も管理してくださっている大畑 遥 氏、井上 祐一郎 氏、そして研究室メンバーに感謝申し上げます。

最後になりましたが、著者が大学、大学院で6年間学ぶことを許していただき、いつも支えてくれた両親、家族、そして親戚一同の方々に感謝の意を申し添えます。

参考文献

- [1] 山本真行, インフラサウンド・音波帯域による大気リモートセンシング, 平成 23 年度スペースプラズマ研究会, 2012.
- [2] Nobuo Arai, Makiko Iwakuni, Shingo Watada, Yuichi Imanishi, Takahiko Murayama, Mami Nogami, Atmospheric boundary waves excited by the tsunami generation related to the 2011 great Tohoku-Oki earthquake, 2011.
- [3] 西山好則, ” 新方式インフラサウンドセンサの開発, ” 高知工科大学 2006 年度学士課程卒業論文, 2007.
- [4] 山田龍樹, ” インフラサウンドによる微小圧力変動の検出方式検討と空振計の開発, ” 高知工科大学 2008 年度学士課程卒業論文, 2009.
- [5] インフラサウンド観測ネットワークについて, <http://infrasound.kochi-tech.ac.jp/infrasound/>, 2019 年 12 月 24 日 参照.
- [6] 反町玲聖, ” インフラサウンド N 型波形イベント自動検出ソフトウェアの開発, ” 高知工科大学 2015 年度学士課程卒業論文, 2016.
- [7] 気象庁, 各種の火山観測 (常時観測・機動観測), <https://www.data.jma.go.jp/svd/vois/data/tokyo/STOCK/kaisetsu/volmonita/volmonita.html>, 2020 年 2 月 2 日 参照.
- [8] 気象庁, 桜島 観測点配置図, https://www.data.jma.go.jp/svd/vois/data/fukuoka/506_Sakurajima/506_Obs_points.html, 2020 年 2 月 2 日 参照.
- [9] 国土交通省 国土地理院, 鹿児島県の山, <https://www.gsi.go.jp/kihonjohochousa/kihonjohochousa41186.html>, 2020 年 2 月 2 日 参照.
- [10] 気象庁, 桜島で発表した火山噴火に関する火山観測報, <https://www.data.jma.go.jp/svd/vois/data/tokyo/STOCK/volinfo/volinfo.php?info=VG&id=506>, , 2019 年 12 月 24 日 参照.
- [11] Christopher, J. Shallue, Andrew Vanderburg, Identifying Exoplanets with Deep Learning: A Five Planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90, 2017.
- [12] Sadaoki Furui, On the role of spectral transition for speech perception, 1986.
- [13] I. Khait, O.Lewin-Epstein R.Sharon, K.Saban, R. Perelman, A. Boonman, Y.Yovel, L.Hadany, Plants emit informative airborne sounds under stress, 2018.
- [14] 株式会社サヤ, インフラサウンドセンサー ADX II INF01 USER MANUAL, https://www.resona-ales.com/_userdata/infrs2.pdf, 2020 年 2 月 23 日 参照.
- [15] 株式会社サヤ, INFRA-SOUND-SENSOR ADXIII-INF01LE user manual, https://www.saya-net.com/_userdata/adx3-inf01le.pdf, 2020 年 1 月 6 日 参照.
- [16] 高知工科大学, 複合型インフラサウンド津波センサーの開発, [57](https://www.kochi-</div><div data-bbox=)

- tech.ac.jp/power/business/post_36.html, 2020年2月23日参照.
- [17] Jrmie Du Boisberranger, Joris Van den Bossche, Loc Estve, Thomas J Fan, Alexandre Gramfort, Olivier Grisel, Yaroslav Halchenko. Nicolas Hug, Adrin Jalali, Guillaume Lemaitre, Jan Hendrik Metzen. Andreas Mueller, Vlad Niculae. Joel Nothman. Hanmin Qin, Bertrand Thirion, Tom Dupr la Tour, Gael Varoquaux, Nelle Varoquaux, Roman Yurchak, scikit-learn, <https://scikit-learn.org/stable/index.html>, 2020年1月12日参照.
 - [18] Andreas C.Muller, Sarah Guido, Python で始める機械学習 scikit-learn で学ぶ特徴量エンジニアリングと機械学習の基礎, 中田秀基 (訳), 株式会社オライリー・ジャパン, 東京, 2017.
 - [19] scikit-learn developers, 3.1. Cross-validation: evaluating estimator performance, https://scikit-learn.org/stable/modules/cross_validation.html, 2020年2月2日参照.
 - [20] 平井有三, はじめてのパターン認識, 森北出版株式会社, 東京, 2012.
 - [21] C. M. ビショップ, パターン認識と機械学習 下, 元田 浩 (翻訳), 栗田 多喜夫 (翻訳), 樋口 知之 (翻訳), 松本 裕治 (翻訳), 村田 昇 (翻訳), シュプリンガー・ジャパン株式会社, 東京, 2008.
 - [22] 田平誠, ” 火山噴火によって発生するインフラサウンドについて, ” 日本音響学会誌 48 巻 10 号, 1992.
 - [23] 反町玲聖, ” インフラサウンドセンサの面的配置における波源位置推定方法の最適化, ” 高知工科大学 2017 年度修士課程卒業論文, 2018.
 - [24] 北川源四郎, 時系列解析入門, 岩波書店, 東京, 2019.
 - [25] K.Hiroe, フーリエ変換の諸性質, <https://eman-physics.net/math/fourier06.html>, 2020年1月7日参照.
 - [26] 板橋秀一、赤羽誠、石川泰、大河内正明、粕谷英樹、桑原尚夫、田中和世、新田恒雄、矢頭隆、渡辺隆夫, 音声工学, 板橋秀一, 森北出版株式会社, 東京, 2005.
 - [27] L. Rabiner, S. Levinson, A. Rosenberg, J. Wilpon, Speaker-independent recognition of isolated words using clustering techniques, IEEE ASSP-27, 4, pp.336-349, 1979.
 - [28] O’Shaughnessy, Douglas, Speech communication: Human and machine, 1987.
 - [29] Beth Logan, Mel Frequency Cepstral Coefficients for Music Modeling, 2000.
 - [30] 篠田浩一, 音声認識 機械学習プロフェッショナルシリーズ, 講談社, 2017.
 - [31] 西村 義隆, 篠崎 隆宏, 岩野 公司, 古井 貞熙, 重みつきスペクトル特徴量を用いた雑音に頑健な音声認識, 2003.
 - [32] 中川聖一, A Survey on Automatic Speech Recognition, 2002.
 - [33] Miyazawa, 動的特徴量 (デルタパラメータ), <http://speechresearch.fiw-web.net/107.html>, 2020年1月7日参照.

- [34] 平田大祐, ” 減災のための Web 上インフラサウンドデータベースの開発, ” 高知工科大学 2018 年度修士課程卒業論文, 2019.
- [35] James D.Murray, マレー数理生物学入門, 三村 昌泰 (総監修), 瀬野 裕美 (監修), 河内一樹 (監修), 中口 悦史 (監修), 丸善出版, 東京, 2014.
- [36] 高橋大輔, 数値計算, 岩波書店, 東京, 1996.
- [37] Google AI Blog, Launching the Speech Commands Dataset, <https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html>, 2020 年 1 月 10 日 参照.
- [38] 本橋智光, 株式会社ホクソエム, 前処理大全 データ分析のための SQL/R/Python 実践テクニック, 技術評論社, 東京, 2018.

付録

不均衡データにおける正解率

機械学習でモデルを作る際に、学習データが不均衡だと、予測精度が下がることが多い [38]。しかし、正解率の場合だと真ラベルにおけるデータ数 P もしくは N の一方のデータ数が多くなると、見かけ上の正解率が向上する。今回英単語 *bed* の音声データが約 3300 あるのに対して、残り 29 単語の音声データは合計で約 62000 あるため、アンダーサンプリングが行える。英単語 *bed* について、アンダーサンプリングを行った結果を図 5.1 に示す。

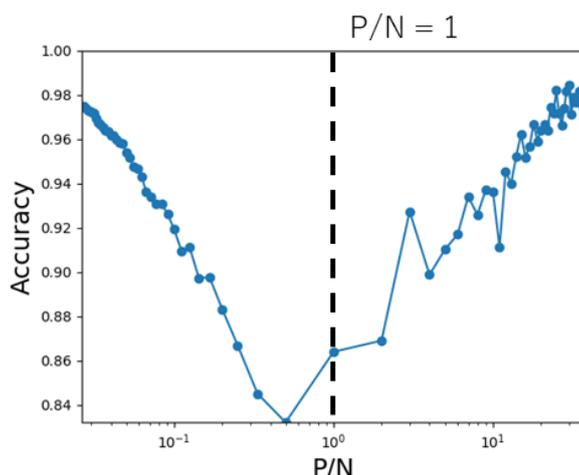


図 5.1 英単語 *bed* によるその他の単語をアンダーサンプリングをした場合における正解率

図 5.1 について、縦軸は正解率であり、横軸は真ラベルのデータ数の比 P/N となっており、横軸については対数プロットである。また $P/N = 1$ の位置を黒破線で表している。 P/N について、アンダーサンプリングのため、初期は $P < N$ であるが、 $P \geq N$ となったときにも N を減少させている。いずれにしても、 $P/N = 1$ から離れていると正解率が向上している。このことは正解率の定義式 (1) において、 $FP + FN \ll TN$ または $FP + FN \ll TP$ であるとき 正解率 ≈ 1 となるためであると考えられる。すなわち、データセットに含まれている真陽性と真偽性の一方の数が偽陽性と偽陰性のデータ数に対して非常に大きくなり、偽陽性および偽陰性のデータ数が正解率に反映されなくなるためである。このためには、不均衡データにならないように $P/N = 1$ 程度になるように、 P 、 N を調整する必要がある。

データセット数に対する再現率と適合率

本文では正解率と F-値のみでデータセット数の変化を評価した。これは F-値が再現率と適合率の調和平均をとっているためである。3.4.2 項では交差検証法によって得られた平均値と標準誤差について、平均値ではピークをもち、データセット数を減らしていくほど標準誤差が小さくなることを示した。この現象は再現率と適合率でも同様の現象が確認でき、データセット数 N を減らした場合において $\mu + \sigma$ による閾値を超えたときの閾値の描画を図 5.2 から図 5.7 に示す。また性能評価値のピークが求められたが、再現率と適合率におけるデータセット数 N_{max} を 5 に示す。さらにデータセット数 N を減らしたときに閾値を超えたときの再現率と適合率における必要データセット数 N_{need} を表 5 に示す。

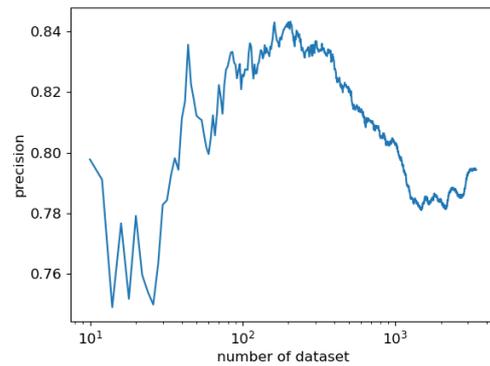
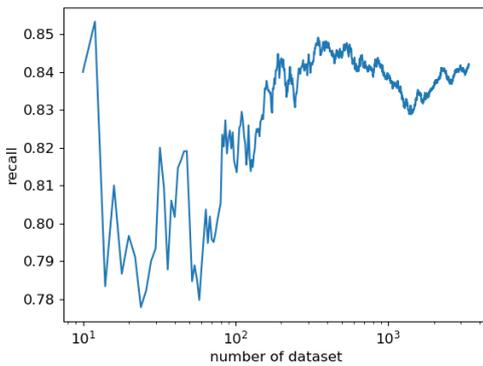


図 5.2 データセット数に対する各英単語の再現率 図 5.3 データセット数に対する各英単語の適合率

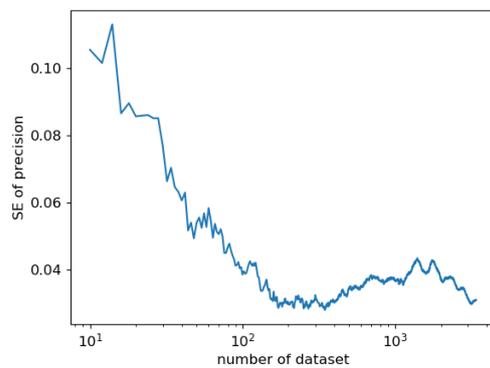
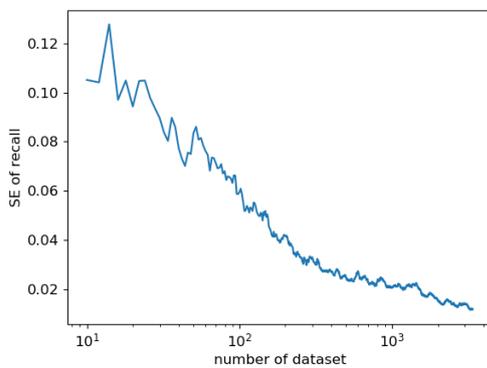


図 5.4 データセット数に対する交差検証法によって得られた再現率の標準誤差

図 5.5 データセット数に対する交差検証法によって得られた適合率の標準誤差

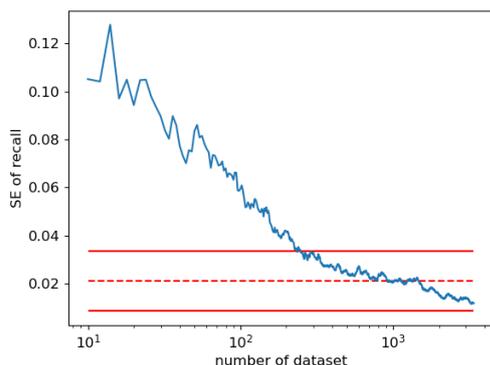


図 5.6 図：データセット数に対する再現率の標準誤差において、閾値による最小データセット数の箇所

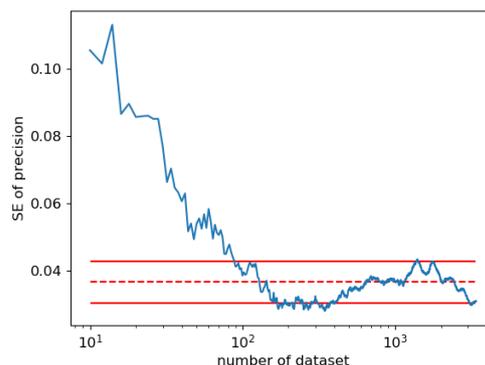


図 5.7 データセット数に対する適合率の標準誤差における閾値による最小データセット数の箇所

表 5.1 データセット数に対する再現率と適合率が最大のときのデータセット数 N_{max}

性能指標名	性能評価値	データセット数 N_{max}
再現率	0.85	348
適合率	0.84	206

表 5.2 データセット数に対する再現率と適合率において交差検証法による標準誤差が閾値を超えたときの必要データセット数 N_{need}

性能指標名	標準誤差	データセット数 N_{need}
再現率	0.03	246
適合率	0.04	88