

## 環境騒音調査の騒音レベルデータを対象とした K 近傍法による異常音判別

大谷 紀子 研究室

1972008 荒川 由人

### 1. 研究の背景・目的

環境基本法第 16 条第 1 項の規定に基づく環境基準は、生活環境を保全し、人の健康の保護に資する上で維持することが望ましい基準であり、大気、水、土壌、騒音をどの程度保つことを目標に施策を実施していくのかを示している。騒音に係る環境基準の達成状況は等価騒音レベルによって評価する。「道路に面する地域」において評価対象となる騒音は道路交通等に起因する騒音であり、環境基準の適用対象外である騒音や、評価の妨げになる騒音は除外しなければならない。除外すべき異常音として、航空機騒音や鉄道騒音、建設作業騒音、鳥の鳴き声などの平常でない自然音、パトカーのサイレンなどの通常は発生しない人工音、測定員の咳払いなどの測定による付加的な音が挙げられる[1]。建設環境研究所は、環境騒音調査において騒音計により騒音レベルデータ、IC レコーダにより音データを記録している。異常音を除外する際には、まず騒音レベルの時刻歴グラフを見て異常音の疑いがある時点を特定する。次に特定した時点の音を聞き、除外を要するかどうかを判別する。異常音の判別作業は、除外作業を担当する人間にとって大きな負担となっている。本研究では、K 近傍法を用い、環境騒音調査における騒音レベルデータから異常音を自動で判別する手法を提案し、異常音の除外作業における作業員の負担を軽減することを目的とする。

### 2. 提案手法

提案手法では、K 近傍法を用いて騒音レベルデータに含まれる異常音を判別する。K 近傍法は、

主に分類に利用される機械学習のアルゴリズムであり、学習データと分類対象データの距離を計算し、近傍の  $k$  個の学習データによる多数決でデータを分類する。騒音レベルデータは、1 次元の時系列データである。したがって、騒音レベルデータを窓幅  $w$  の部分時系列に分け、 $w$  次元ベクトルにしてモデルに入力する。

提案手法における、K 近傍法を用いた異常音判別の手順を以下に記す。

- ① 学習用の騒音レベルデータから、一定の窓幅の部分時系列を取り出し、ベクトル化する。
- ② 判別対象の騒音レベルデータから、学習時と同じ窓幅の部分時系列を取り出し、ベクトル化する。
- ③ 判別対象データと、すべての学習データとの距離を計算する。
- ④ ③で計算した距離が短い順に  $k$  個の学習データを選び、判別対象データとの距離の平均値を異常度とする。
- ⑤ 異常度が閾値を超えているデータを異常として判別する。

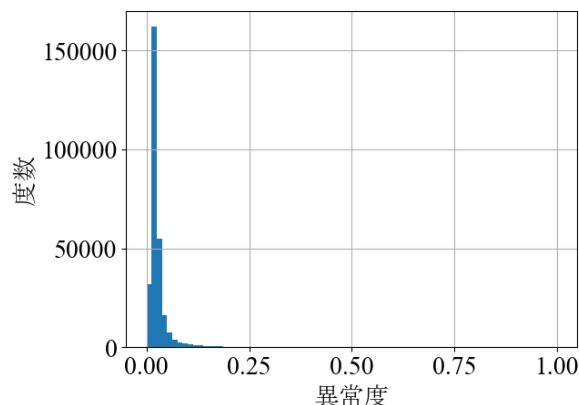


図 1: 異常度のヒストグラム

異常度の適切な閾値は、学習データと判別対象データの組み合わせによって異なるため、データに応じて異常度の閾値を変更する必要がある。

判別対象データの異常度のヒストグラムを図 1 に示す。騒音レベルデータの大部分は正常であるという特徴から、一次元上で正常な異常度データは密集し、異常な異常度データは広範囲に散在する。したがって、閾値設定をヒストグラムの分割と捉えることができる。提案手法では、K 平均法と四分位範囲を用いた外れ値検出法を組み合わせることで、ヒストグラムの分割点を特定し、データに応じて適切な閾値を自動で設定する。K 平均法は、クラスタリングに利用される機械学習アルゴリズムであり、事前に指定されたクラスタ数に基づきデータを分割する。四分位範囲を用いた外れ値検出法は、四分位範囲を 1.5 倍に拡大した範囲から外れるデータを検出する。

提案手法における閾値の設定手順を以下に記す。

- ① K 平均法を用いて、判別対象データの異常度を表す 1 次元データ列を、以下に記す 2 つのクラスタに分割する。
  - A) 密な異常度データ
  - B) 疎な異常度データ
- ② クラスタ B に属する異常度データに対して、四分位範囲を用いた外れ値検出を行う。
- ③ クラスタ B に属する異常度データの外れ値を除外する。
- ④ クラスタ B に属する異常度データの最大値を閾値とする。

### 3. 評価実験

評価実験では、建設環境研究所が用意した 9 日分の騒音レベルデータのうち 8 日分を学習データ、1 日分をテストデータとする 9 分割クロスバリデーションを行い、総合的にモデルの性能を検証する。長時間にわたる騒音レベルデータに正確なラベルを人手で付与するのは非常に困難であるため、教師ラベルは大雑把となっている。例えば、「カラ

スが鳴いている」という大雑把なラベルが付与された区間には、「鳴いている」状態と「鳴き止んでいる」状態が交互に混在しており、本来はそれぞれの状態ごとに細かく切り分けて緻密にラベルを付与する必要がある。

異常度算出モデルの評価指標として、PR 曲線の AUC を用いる。PR 曲線は閾値を変化させながら、適合率と再現率をプロットしたグラフである。適合率は正例と予測したデータのうち、実際に正例であるものの割合であり、再現率は実際に正例であるもののうち、正例であると予測されたデータの割合である。評価実験では、教師ラベルが大雑把であるため、一般的な再現率で評価することができない。したがって、異常ラベルが付与された区間のうち、異常と判別されたデータを含む区間の割合を再現率とする。AUC は PR 曲線の下部の面積を表す。面積が最大で 1、最小で 0 となり、1 に近づくほど精度が高いことを示す。AUC を用いることで、閾値の影響を受けずに、汎用的なモデルの性能を評価することができる。閾値設定手法の評価指標としては適合率、再現率、F 値を用いる。F 値は適合率と再現率のトレードオフ関係に着目し、2 つの値を調和平均した評価指標である。

### 4. 結果と考察

評価実験の結果、AUC は 0.921、F 値は 0.838 となり、提案手法が優れた異常判別能力を持つことがわかった。しかし、K 近傍法を用いた異常判別モデルは、すべてのデータ点の距離を考慮し、低密度なクラスに属するデータを異常とするため、バイクの走行音など“正常だが希少な音”の異常度が高くなるという問題がある。今後の課題として、学習データに含まれる“正常だが希少な音”をオーバーサンプリングするといった改善が必要である。

### 参考文献

- [1] 環境省, "騒音に係る環境基準の評価マニュアル 道路に面する地域編", 2015.