

環境音を用いた雰囲気の可視化方式の検討

Study for Visualization Method of Area Feature using Environment Sound

小野 祐介

Yusuke Ono

岡山大学 阿部研究室

Abe Laboratory, Okayama University

概要 本研究では、身の回りの環境に依存する音情報（環境音）を用いて地域特性を分類し、雰囲気が想起できる形で可視化する方式を検討した。本稿では、すでに収録済みの環境音データを用い、場所を観光地、繁華街、住宅地の3つの地域特性に分類し、地図上に可視化した。使用する特徴量は、音の大きさ、高さ、音源の種類、収録時間帯であり、それぞれの特徴量の重要度を分析した。分析結果をもとに新たな特徴量を追加し、分類精度を比較した。

1 はじめに

音は身の回りの状況を理解するための重要な情報の一つである。特に、人間の生活を取り巻く環境音は、人の話し声や車の音などの多種多様な音源だけでなく、その場の状況や雰囲気など、多くの情報を含んでいる。

環境音を分析して、その中に含まれた情報を取り出し、活用する研究が行われている。以下に具体例をあげる。

1. 騒音度を地図上に可視化した騒音マップを用いた騒音問題発生の抑制
2. 賑わい度を地図上に可視化することで観光事業や飲食業などの営業形態支援
3. 雰囲気を想起させる特徴を地図上に可視化し、場所の雰囲気理解を支援

例えば、環境音のうち自動車から発せられる騒音を可視化する方法として、国立騒音研究所による自動車常時監視をもとにした全国自動車交通騒音マップ [1] がある。全国自動車交通騒音マップ [1] では、騒音の大きさをもとにしているが、人によって騒音の感じ方が異なることが知られており、それを踏まえて可視化する研究もなされている [2]。文献 [2] では、スマートフォンやタブレットにインストールされた”オトログマップ” [3] や”オトログ!” [4] というアプリケーションを用いて実地で収録された環境音データから人間が感じる騒音度を推定し、騒音マップとして可視化する研究をしている。他にも、多くの人に各個人の所有するスマートフォンを用いて環境音の収集を行う、クラウドソーシングの方式で得られた環境音データを用いて、騒音度を可視化する研究が行われている [5]。

文献 [6] では、商業施設の営業時間を WEB 上で API

を用いて取得し、緯度経度や電話番号などの住所を用いて地図上に1時間ごとに店の位置を可視化している。また、GPS ログを用いて得られた人口密度データを可視化している。

また環境音を、生物によって生成された音 (biophony)、人間によって生み出される音 (anthrophony)、雨、風などの非生物による音 (geophony)、の3種に分類し、時間的、空間的な音響パターンをサウンドスケープエコロジーと定義している [7]。文献 [8] では、これらの分類をもとに都市で聞こえる人為的な音（車の走行音など）を音楽でマスクし、それ以外の音（鳥の声や、風の音など）に注意を向けさせるアプリケーションの作成をしている。この論文では、雰囲気を人為的でない音としてとらえている。また、3D 仮想風景と環境音を組み合わせることで、その場にいる感覚を味わえることができるかを評価している研究がある [9]。

本研究では、雰囲気を想起させる特徴を地図上に可視化し、場所の雰囲気理解を支援をする手法を検討する。

2 提案手法

雰囲気理解を支援するために必要な情報として、”地域ごとに人が何の目的で訪れるのか”が1つあげられる。人が活動している地域は、以下の3つの地域特性に分類できると考えた。

1. 住宅街
2. 繁華街
3. 観光地

住宅街とは、人が居住している地域のことであり、繁華街とは、商業施設が多く集まる地域のことであり、観光地とは、歴史、文化、自然景観などが整備されており、交通機関や宿泊施設などで観光客の受け入れを行える地域のことであり、例えば、図1のように可視化する。

可視化を行うにあたって、今後収録された未知の環境音データを用いて、地域特性を自動的に分類することが望ましいと考えられる。そのため、環境音データを用いて、地域特性を分類するモデルの作成を行う。

2.1 使用した環境音データ

環境音データは、Android 端末で動作するアプリケーション、オトログマップ [3] やオトログ! [4] を用

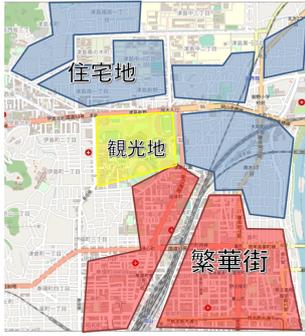
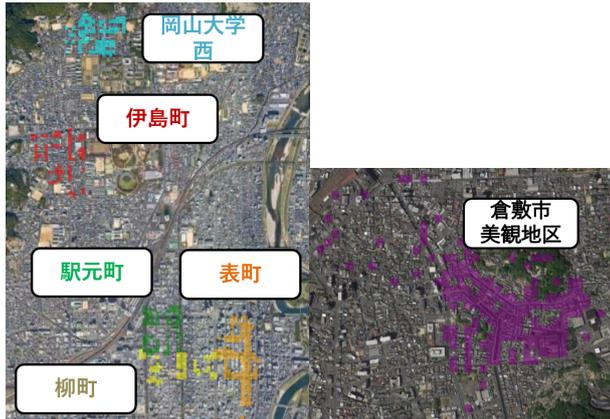


図1 地域特性を3つに分類(青:住宅街, 赤:繁華街, 黄:観光地)



(a) 岡山駅周辺 (b) 倉敷市美観地区周辺

図2 収録場所の詳細

いて、実地で収録されたデータを用いている。その場で感じる主観的騒音レベルやその場で聞こえた環境音の種類を入力し送信すると同時に、直近10秒または15秒の音声波形をwavファイルとして残し、それに対応する収録日時と緯度経度のログファイル、音声データから計算された人間の聴覚特性に基づく等価騒音レベル (L_{Aeq})、各周波数帯 (62.5 Hz, 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz) のパワーを三角窓を用いて近似的に求めたバンドパスフィルタ出力 (BP) のログファイル、主観的騒音レベルなどの入力された情報のログファイルを記録している。

環境音収録は岡山県岡山市、および、同県倉敷市で行っている [3, 4]。収録場所の詳細を図2に示す。

また、環境音の収録状況を表1に示す。使用した機器は、2014年、2015年収録のものはGoogle Nexus7、2016年、2017年収録のものはそれに加えてスマートフォン (SH-01F など) を用いている。スマートデバイスのマイクロホン特性については校正済みである。

環境音データの内容を表2に示す。

環境音ラベルとは、その場で聞こえた環境音の種類である。 T_{01} : 人の声, T_{02} : 鳥の声, T_{03} : 虫の音, T_{04} : 車の音, T_{05} : 風切音, T_{06} : バイクの音, T_{07} : 踏切の音, T_{08} : 電車の音, T_{09} : サイレン音, T_{10} : 音響信号機

表1 環境音収録状況

収録日	収録場所	収録参加者人数
2014年11月	岡山大学周辺	
2015年1月	岡山駅周辺	14
2016年5,10月	倉敷市美観地区	35
2017年9,10月	倉敷市美観地区	38

表2 使用した環境音データの内容

area_feature	地域特性
area_id	収録場所
device_id	収録デバイス
timestamp	収録日時
quadkey	位置情報
sound_label	音源の種類
L_{Aeq}	等価騒音レベル
BP	バンドパスフィルタ出力

の音, T_{11} : 音楽, T_{12} : 動物の声 (鳥を除く), の12種類の環境音の音源をラベル付けしている。10秒または15秒の間に聞こえた音であれば、いくつでもラベルを選んでよいとしている。

2.2 地域特性の分類

収録した地域は, "住宅地", "繁華街", "観光地" の3つの地域特性に大別できることをある程度想定して収録されている。そのため, 収録場所を用いて地域特性の正解データを与えることにした。岡山大学西, 伊島町を住宅地, 駅元町, 柳町, 表町を繁華街, 倉敷市美観地区を観光地とした。

2.3 データの前処理

まず, 表2で示した特徴量のうち, 場所に関する情報を用いてはならないことに注意する。理由は, 地域特性の分類を場所をもとに行っているからである。よって, 使用する特徴量は, 以下の通りになる。合計で62次元の特徴量となった。

- 等価騒音レベル (L_{Aeq})
 - 10 sec の音声波形データの1 sec ごとの L_{Aeq} ($leq[9-0]$)
 - $leq[9-0]$ の平均 (avg_leq)
 - 各 sec の L_{Aeq} の10パーセンタイル ($pleq-10$), 50パーセンタイル ($pleq-50$), 90パーセンタイル ($pleq-90$)
- バンドパスフィルタ出力 (BP) : 各周波数帯 (62.5 Hz, 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz) での10 sec の音声波形データの1 sec ごとのBP ($bp[7-0]-[9-0]$)
 - 各周波数帯のBPの平均 ($avg_bp[7-0]$)
 - 各 sec の L_{Aeq} の10パーセンタイル

表3 ベースラインモデルの推定精度

	precision	recall	f1-score	support
住宅地	0.777	0.691	0.732	2089
繁華街	0.761	0.852	0.804	3626
観光地	0.909	0.879	0.894	6238

(pbp[7-0]_10), 50 パーセントイル (pbp[7-0]_50), 90 パーセントイル (pbp[7-0]_90)

- 音源ラベル (T[01-12])
- 時間帯 (朝: time_before_11, 昼: time_after_16, 夜: time_on_11_16)

音源ラベルは、該当した場合を1、それ以外を0とする 1-of-K 表現を用いた。時間帯に関しては、朝 (収録開始時刻から 11 時 00 分まで)、昼 (11 時 00 分から 17 時 00 分まで)、夜 (17 時から収録終了時刻まで) に分割し、1-of-K 表現を用いた。等価騒音レベルとバンドパスフィルタ出力については、最小値が 0、最大値が 1 となるように正規化している。

2.4 推定器の構築

木構造モデルのランダムフォレスト分類器を用いて、地域特性の分類器を学習した。実装には python 3.8.3 上で scikit-learn 0.22.1 を用いた。ランダムフォレストの決定木数 (n_estimators) を 1000 とし、スコアリング関数は負の対数尤度を使用している。

また、分類器の過学習を抑制するために、グリッドサーチによる最適な分割閾値 (min_sample_split) の探索と、グリッドサーチと学習それぞれに 5 分割交差検証を行っている。

2.5 各特徴量の重要度

ランダムフォレスト分類器を用いて分類するにあたって、効果的な特徴量を分析するために、各特徴量の重要度を計算した。

等価騒音レベルとバンドパスフィルタ出力の 10 パーセントイルの重要度が高くなっており、音源ラベルでは、車の音、音響信号機、人の声の重要度が高くなっていった。

3 実験結果

3.1 ベースライン

まずは、2.3 節で述べた特徴量を用いて、分類器を作成した。推定精度を表 3 に示す。

3.2 考察

表 3 の結果を見ると、観光地、繁華街、住宅地となるにつれて推定精度が低くなっていることがわかる。理由の一つとして、データ数に偏りがあるからではないかと考えられる。観光地のデータ数は、6278 であり、繁華街は 3626、住宅地は 2089 である。次に、データ

表4 データ数を平滑化したモデルの推定精度

	precision	recall	f1-score	support
住宅地	0.811	0.778	0.794	2089
繁華街	0.742	0.844	0.790	2089
観光地	0.879	0.794	0.835	2089

表5 新たな特徴量を追加したモデルの推定精度

	precision	recall	f1-score	support
住宅地	0.796	0.688	0.738	2089
繁華街	0.775	0.863	0.817	3626
観光地	0.910	0.892	0.901	6278

数をすべて等しくして推定を行った。

3.3 データ数平滑化

住宅地のデータ数 2089 に、繁華街、観光地のデータ数を揃えた。データはランダムに選んで揃えている。推定精度を表 4 に示す。

3.4 考察

表 4 を見ると、データ数にばらつきがある時と比べて、住宅地の推定精度が上がっていることがわかる。ただし、繁華街と観光地に関しては推定精度が下がっている。このことから、データ数の割合に応じて推定精度が左右されるのではないかと考えられる。

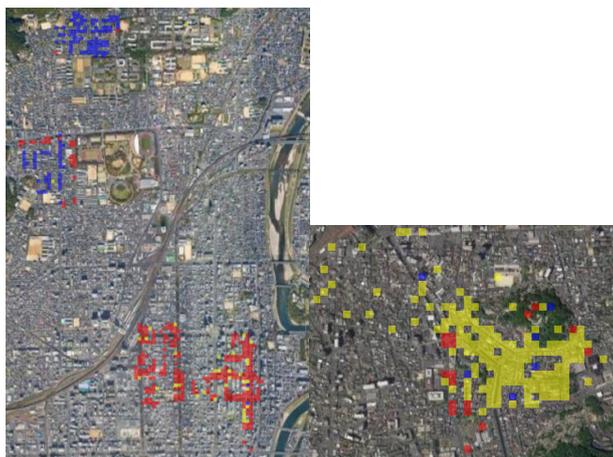
3.5 新たな特徴量の追加

2.5 節で特徴量の重要度を分析し、等価騒音レベルと各周波数帯のバンドパスフィルタ出力での 10 パーセントイルが特徴量として重要であることが分かった。しかし、現時点では 10 または 15 秒の音声波形データから計算された値を用いているため、時間的に収録場所の地域特性を十分に表現できていないのではないかと考えられる。そこで、収録時刻 (timestamp) から前後 30 分までで同一の場所 (quadkey) で計測された等価騒音レベルの平均を新たに計算し、特徴量として用いることにする。全データを用いて推定を行った結果のうち推定精度を表 5 に示し、推定結果を図 3 に示す。同一の場所で複数の推定結果があるところでは、推定結果の最頻値を表示するようにしている。

3.6 考察

表 2 と比べると、それぞれ F1-score が、住宅地では 0.006 ポイント、繁華街では 0.013 ポイント、観光地では 0.007 ポイント上昇した。

各特徴量の重要度を分析した結果を図 4 に示す。赤枠で示した提案した特徴量の重要度が全 62 次元の特徴量のうち、8 番目であった。このことから、追加した特徴量は今回の分類タスクにおいて有効であると考えられる。



(a) 岡山駅周辺 (b) 倉敷市美観地区周辺

図3 新たな特徴量を追加したモデルの推定結果

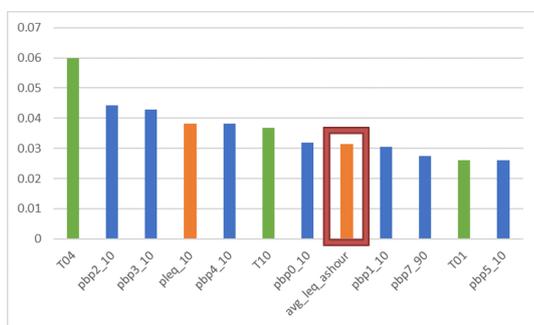


図4 提案した特徴量（赤枠）を含めた各特徴量の重要度

4 まとめと今後の課題

本研究では、すでに収録済みの環境音データ [3, 10, 4] から、住宅地、繁華街、観光地の3つの地域特性をランダムフォレスト分類器を用いて推定し、結果を地図上に可視化した。3分類だけでなく、閑静か賑わっているかなどを追加で分類し、より詳細な地域特性情報を生成することを検討していく。また、推定精度にデータ数の偏りが影響していることから、少ないデータ数の1つのデータの最近傍データを用いて新しいデータを生成するSMOTEなどのアルゴリズムを用いてデータ数を均一化する必要があると考えられる。また、新たに追加した特徴量を用いて推定を行った結果、分類器の精度が上がり、特徴量の重要度も高いことから新たに追加した特徴量は有効だと考えられる。今後は、推定精度を向上させるだけでなく、誤った推定結果であってもその地域の新たな特性を表現している可能性を考慮し、より有効な雰囲気可視化方式を検討する。

参考文献

- [1] “国立環境研究所：自動車騒音の常時監視結果,” <http://tenbou.nies.go.jp/>, accessed July. 2020.
- [2] S. Kobayashi, M. Abe, and S. Hara, “Prediction

of subjective assessments for a noise map using deep neural networks,” Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp.113–116, 2017.

- [3] S. Hara, S. Kobayashi, and M. Abe, “Sound collection systems using a crowdsourcing approach to construct sound map based on subjective evaluation,” 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)IEEE, pp.1–6 2016.
- [4] 朝田興平, 原直, 阿部匡伸, “クラウドソーシングによる賑わい音識別方式のフィールド実験評価,” 2018年日本音響学会春季研究発表会講演論文集, no.1-Q-7, pp.79–82, March 2018.
- [5] N. Maisonneuve, M. Stevens, M.E. Niessen, and L. Steels, “Noisetube: Measuring and mapping noise pollution with mobile phones,” Information technologies in environmental engineering, pp.215–228, Springer, 2009.
- [6] 岡本裕紀, “営業時間から見た商業集積地の賑わい変化の推定,” 2013.
- [7] B.C. Pijanowski, L.J. Villanueva-Rivera, S.L. Dumyahn, A. Farina, B.L. Krause, B.M. Napolitano, S.H. Gage, and N. Pieretti, “Soundscape ecology: the science of sound in the landscape,” BioScience, vol.61, no.3, pp.203–216, 2011.
- [8] T. Carson, “Sounds aware: A mobile app for raising awareness of environmental sound”. <https://www.tatecarson.com/>, accessed July. 2020.
- [9] M. Lindquist, E. Lange, and J. Kang, “From 3d landscape visualization to environmental simulation: The contribution of sound to the perception of virtual environments,” Landscape and Urban Planning, vol.148, pp.216–231, 2016.
- [10] S. Hara, A. Hatakeyama, S. Kobayashi, and M. Abe, “Sound sensing using smartphones as a crowdsourcing approach,” 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)IEEE, pp.1328–1333 2017.