

クラウドソーシングによる環境音マップ構築のための 主観的な騒々しさ推定方式の検討

Prediction of subjective loudness levels to construct environmental sound map with crowdsourcing approach

原直[†]

Sunao Hara

阿部 匡伸[†]

Masanobu Abe

1. はじめに

音は人間にとって周囲環境を理解するための重要な手段の一つであり、音を基盤としたセンシング技術はスマートシティにおける環境理解のための重要な技術となりえる。スマートシティにおけるデータ収集は継続的かつ詳細に収集することが望ましく、このようなデータ収集の実現には住民による協力を得ることが不可欠である [1]。近年の携帯電話やスマートフォンは多様なセンサー群を搭載しているものの、端末によって搭載されるセンサーの種類が異なることも多いが、GPS (Global Positioning System) とマイクロホンほとんどの端末に内蔵されている。このことから、住民が持つスマートホンを用いて位置情報を付与した音を収録するためのハードウェア環境は整いつつあると言える。

本研究ではクラウドソーシングによるデータ収集方式として知られる便乗型センシング (Opportunistic sensing [2]) と参加型センシング (Participatory sensing [3]) という 2 つの考え方に基づいた Android 用アプリケーションを作成し、データ収集を進めている [4]。便乗型センシングとしてはアプリケーション起動中は常にその場の騒音レベルを記録しており、音の物理的な特徴を表す統計量の一つとして騒音レベルを収集することができる。一方、参加型センシングとしては聴取者が気になった音を記録することを可能としており、これによって音声波形そのものとその場における主観的な印象を収集することができる。

環境音の可視化は音による環境理解に向けた重要な機能の一つである。本研究では、位置情報を持つ環境音の情報を地図上で可視化した地図のことを環境音マップと呼んでいる。これまでも、鳥の声などが聞こえた地点へのアイコン表示や、特定区画毎の等価騒音レベル統計量を表示した騒音マップ [5, 6, 7] を環境音マップとして用いてきた [4, 8]。しかし、音の大きさのみで作成した騒音マップと主観的な感じ方に基づいた騒音マップで

は、差が見られた [9]。

本報告では、音の大きさだけを基本的な情報源としながら、補助的な情報を加えることで、音の大きさの感じ方を推定する方式について報告をおこなう。補助的な特徴量としては、オクターブバンドフィルタ出力、場所、日時、音源を加える。音の大きさの感じ方についての正解データとして、データ収録者が現地で感じて入力した 5 段階の騒音度合いを用いる。特徴量の組み合わせ方や推定器の種類による性能の変化について実験をおこない、その性能評価をおこなう。

2. 環境音データベース

実環境で収録された環境音データベース [4, 8, 10] を用いる。本章ではデータベースの収録に使われたアプリケーションの概要と収録データの概要について述べる。

2.1 環境音収録アプリケーション

データ収録は Android 端末で動作するアプリケーションで行われている。音を保存する際にはその場の音を主観的に評価するために騒音度と混雑度の 5 段階評価を与える事ができる。さらに、その場の状況や聞こえてくる音を記録するために、任意入力のテキストエリアと 12 種類のあらかじめ設定された音の種類が用意されている。アプリケーション内の送信ボタンが押されると、直近 10 秒または 15 秒の音声波形を WAV ファイルとして作成すると同時に、主観評価の選択項目とテキスト入力の内容を表す文字列に時間をつけてログファイルに記録する。

環境音は量子化ビット数 16 bit、サンプリングレート 32,000 Hz で収録が行われる。1 秒毎に得られる環境音サンプルに対して A 特性に基づく等価騒音レベル LAeq を計算する。なお、端末毎のマイクロホンの收音特性の異なりを補正するため、騒音計 (RION NL-42) を用いて線形補間による校正関数を組み込んでいる。

主観的な騒音度合いの選択肢は、 L_1 : とても静か、 L_2 : 比較的静か、 L_3 : やや騒がしい、 L_4 : かなり騒々しい、 L_5 : とても騒々しい、の 5 段階である。そして、環境音

[†] 岡山大学

の種類としては、 T_1 : 人の声、 T_2 : 鳥の声、 T_3 : 虫の音、 T_4 : 車の音、 T_5 : 風切音、 T_6 : バイクの音、 T_7 : 踏切の音、 T_8 : 電車の音、 T_9 : サイレン音、 T_{10} : 音響信号機、 T_{11} : 音楽、 T_{12} : 動物の声 (鳥を除く)、の 12 種類を用意し、複数選択が可能である。

2.2 収録データの概要

環境音収録は岡山県岡山市、および、同県倉敷市で行った。岡山市では、2014 年 11 月と 2015 年 1 月にのべ 14 名で収録をおこなった [4]。岡山駅前の賑わった環境だけではなく、商店街や住宅街などの比較的静かな地域で収録している。倉敷市では、静謐な観光地である倉敷美観地区において、2016 年 5 月と 10 月にのべ 35 名 [8]、2017 年 9 月と 10 月にのべ 38 名 [10] の規模で収録している。特に、2016 年 10 月と 2017 年 10 月の収録日には同地区内にある倉敷阿智神社を中心とした秋季例大祭が催されており、特に多くの観光客が訪れていた。

実験に用いるデータ数は合計で 11,993 個である。主観的な騒音度合いごとの個数としては、 L_1 から L_5 の 5 段階について、それぞれ、1,269 個、4,440 個、4,635 個、1,356 個、293 個となっている。

3. 推定に用いる特徴量

主観的な騒音度合いを推定するために、音の大きさだけでなく、多様な情報を組み合わせた推定を行う。なお、いずれの特徴量についても、学習データの最小値が 0 かつ最大値が 1 となるように正規化している。

3.1 等価騒音レベル LAeq

音の大きさを示す指標であり、主観的な騒音度合いに関する最も基本となる特徴量である。本稿では、主観的な騒音度合いが記録された際の直近 10 秒の等価騒音レベルから、統計量として、平均 (LAeq-avg)、10 パーセンタイル値 (LAeq-p10)、50 パーセンタイル値 (中央値; LAeq-p50)、90 パーセンタイル値 (LAeq-p90) を計算し、4 次元の特徴量として利用する。

3.2 オクターブバンドフィルタ出力 BP

音の高さ成分ごとの成分の強さを表す指標である。収録アプリケーションでは三角窓を用いた近似的なオクターブバンドフィルタの出力値を得ている。中心周波数は 62.5 Hz、125 Hz、250 Hz、500 Hz、1,000 Hz、2,000 Hz、4,000 Hz、8,000 Hz であり、それぞれの中心周波数をもつバンドパスフィルタからの出力を BP1、BP2、…、BP8 とする。等価騒音レベルと同様の方式で各出力ごとに統計量を計算することにより、合計で 32

次元の特徴量を得る。例えば、BP1-avg や BP4-p10 等である。

3.3 場所情報 LC

QuadKey^{*3}により量子化された収録地点であり、そのエリア ID を特徴量として用いる。符号長 13 の QuadKey^{*4}を用いることで、実験に用いるデータは岡山県岡山市 2 エリア、岡山県倉敷市 4 エリアに分割されている。1-of-K 表現により 6 次元の特徴量とした。

3.4 日時情報 DT

収録された日の情報と、収録された時刻に関する情報を、特徴量として用いる。

まず、収録された時刻をおおよその時間帯に分割することで、特徴量として用いる。本研究では気象庁による 1 日の時間区分を参考にして、朝 (10 時 59 分以前)、昼 (11 時 00 分から 16 時 59 分まで)、晩 (17 時 00 分以降) の 3 つの時間帯に分割した。昼を切片として除外したダミー変数とすることで、2 次元の特徴量とした。

さらに、収録された日の属性に基づいてクラスタリングをすることで、特徴量として用いる。土曜日・日曜日・祝日であるか否か、の 1 次元と、その日にイベントが開催されているか否か、の 1 次元からなる、合計 2 次元の特徴量とした。

3.5 音源情報 SS

12 種類の音源ラベルが主観的な騒音度合いと共に付与されていることから、このラベルを特徴量として用いる。各音源ラベルの有無を表す 1-of-K 表現により、12 次元の特徴量とした。

4. 主観的な騒音度合いの推定

L_1 から L_5 という 5 クラスの推定問題として、主観的な騒音度合いの推定をおこなう。ベースラインとなる特徴量は LAeq とする。この特徴量に 3 節に述べた特徴量を順次加えながら、推定精度の変化を調査する。

推定精度の計算は 10-fold 交差検証による。推定対象は人手でつけられた評価値であり、個人ごとの評価値の偏りが存在すると考えられる。そこで、評価者による評価値のバイアスを無視するため、10 folds に分割する際には、各評価者は 1 つの fold にしか存在しない、という制約を加えた。これにより、学習用データには存在し

^{*3} Bing Maps Tile System: <https://msdn.microsoft.com/library/bb259689.aspx>

^{*4} 岡山の緯度 34.6 度付近では一辺がおおよそ 4026.8 m の正方形が 1 エリアに相当する

ない評価者が与えた評価値を推定する問題として定式化した。

4.1 推定器の構築

本稿では、推定精度の比較のために、線形モデルのロジスティック回帰分類器と、木構造モデルのランダムフォレスト分類器を用いて、2種類の推定器を作成した。ロジスティック回帰分類器については、one-vs-rest (one-vs-allとも呼ぶ)によって5クラスの推定器を構築した。これは各クラスごとにその有無を判定する識別器を作り、計5つを組み合わせる方式である。実装にはpython 3.6.5上でscikit-learn 0.19.0を用いた。

分類器の過学習を抑制するために、ハイパーパラメータの一部はグリッドサーチにより最適パラメータを探索した。グリッドサーチでは学習データの一部を検証用データとする5-fold交差検証をおこなっている。対象となるハイパーパラメータは、ロジスティック回帰では正規化強度C、ならびに、ランダムフォレスト分類器の分割閾値min_samples_splitである。

学習時にはクラスごとのデータ数の不均衡による推定モデルの劣化を防ぐため、クラスごとの重みを与えている。事前実験において、この重み付与操作をしない場合、クラス L_1 やクラス L_5 のようなデータ数の少ないクラスが全く推定結果に表れないような識別器が作られる傾向があったが、重み付与をすることで偏った傾向が抑制されることを確認している。

4.2 評価指標

推定器の精度は以下のように計算される。主観的な騒音度合い c の分類器について、再現率(recall) $r^{(c)}$ 、適合率(precision) $p^{(c)}$ 、F1 score $f^{(c)}$ を計算する。

$$r^{(c)} = \frac{tp^{(c)}}{tp^{(c)} + tn^{(c)}},$$

$$p^{(c)} = \frac{tp^{(c)}}{tp^{(c)} + fp^{(c)}},$$

$$f^{(c)} = \frac{2 \cdot p^{(c)} \cdot r^{(c)}}{p^{(c)} + r^{(c)}}.$$

そして、主観的な騒音度合い毎のデータ数 $n^{(c)} = tp^{(c)} + tn^{(c)}$ に依存した重み付き平均を計算し、評価指標として用いる。なお、重み付け操作のため、F1 scoreはRecallとPrecisionの間の値にならない場合がある。

$$\text{Recall} = \frac{\sum_{\forall c} r^{(c)} \cdot n^{(c)}}{\sum_{\forall c} n^{(c)}}, \quad (1)$$

$$\text{Precision} = \frac{\sum_{\forall c} p^{(c)} \cdot n^{(c)}}{\sum_{\forall c} n^{(c)}}, \quad (2)$$

$$\text{F1 score} = \frac{\sum_{\forall c} f^{(c)} \cdot n^{(c)}}{\sum_{\forall c} n^{(c)}}. \quad (3)$$

4.2.1 推定結果

主観的な騒音度合いの推定結果を表1に示す。 F_1 から F_4 へと特徴量を増やすにつれて、評価指標が向上する傾向が見られる。変化量に注目してみると、BPを追加した場合(F_1 から F_2)やSSを追加した場合(F_3 から F_4)に、大きく性能が向上している。このことから、音の大きさだけではなく、音の高さに関する情報や、その場で聞こえた音の種類なども、騒音度合いの推定に寄与しているといえる。

分類器ごとの推定結果に注目してみると、ロジスティック回帰分類器ではF1 scoreの最大値が0.487となり、ランダムフォレスト分類器ではF1 scoreの最大値は0.493となった。わずかではあるが、ランダムフォレスト分類器の方が精度の高いが、F1 scoreがRecallやPrecisionに比べて低い値を示していることから、騒音度合いごとの推定結果には偏りがある可能性が示唆される。

そこで、騒音度合いごとの識別結果を並べた表2を用いてさらに考察を行う。この表では、2つの分類器とベースラインとなる特徴量 F_1 と最大性能を示した特徴量 F_4 の組み合わせからなる4つの推定器の結果のみを示している。騒音度合い L_4 の評価指標に注目すると、ランダムフォレスト分類器では F_1 よりも F_4 を用いた場

表1 主観的な騒音度合いの推定結果

(a) Logistic regression			
Features	Recall	Precision	F1 score
F_1 : LAeq + BP	0.403	0.450	0.420
F_2 : F_1 + LC	0.445	0.453	0.442
F_3 : F_2 + DT	0.455	0.441	0.435
F_4 : F_3 + SS	0.497	0.485	0.487
(b) Random forest			
Features	Recall	Precision	F1-score
F_1 : LAeq + BP	0.411	0.424	0.416
F_2 : F_1 + LC	0.469	0.493	0.458
F_3 : F_2 + DT	0.484	0.506	0.470
F_4 : F_3 + SS	0.511	0.522	0.493

表2 各推定器におけるクラスごとのF1 score

Level	Logistic regression		Random forest	
	F_1	F_4	F_1	F_4
L_1	0.279	0.508	0.237	0.398
L_2	0.478	0.518	0.474	0.562
L_3	0.538	0.535	0.496	0.582
L_4	0.000	0.262	0.177	0.106
L_5	0.208	0.222	0.161	0.229

合の性能が下がっている。一方で、ロジスティック回帰分類器では特徴量 F_1 よりも F_4 を用いた場合の方が高い性能を示している。また、騒音度合い L_1 で両分類器の性能を比較すると、ロジスティック回帰分類器のほうがランダムフォレスト分類器よりも優れた性能を示している。ただし、データ数が多い L_3 や L_4 ではランダムフォレスト分類器の方が高い性能を示している。このことから、3 節に挙げた全ての特徴量が利用できる場合には、ランダムフォレスト分類器よりロジスティック回帰分類器を用いたほうが、データ数がアンバランスでも騒音度合いの頑健な推定が可能になることが示唆される。

5. おわりに

本稿では、收音者が收音中に感じていた主観的な騒音度合いを、音の大きさと様々な情報源を組み合わせながら推定する実験を行った。実験では收音が付与した5段階の騒音度合いを正解値とし、特徴量に等価騒音レベルといくつかの補助情報を用いて、推定器を学習した。実験の結果、ランダムフォレスト分類器によりF1 scoreで約0.49の推定精度が得られた。一方、騒音度合いごとの推定精度に注目してみると、ランダムフォレスト分類器では特定の度合いの推定精度が低くなる現象が見られたが、ロジスティック回帰分類器ではデータ数に依らない均一な推定精度が得られることが示された。

今後の課題としては、実際に地図上に可視化した際の、推定精度の多寡が与える影響について検討する必要がある。また、Deep Neural Networkなどのより高度なモデルを用いる[9]ことで、精度の向上が期待される。さらに、本稿では音源情報は収録者が手動で選択した情報を利用したが、近年の環境音からの音イベント検出(例えば,[11])で用いられる技術を応用することで、音源情報の入力は不要となり、音のみから自動的に騒音度合いが推定できるようになる可能性も考えられる。

謝辞

本研究はJSPS 科研費18K02862の助成を受けて実施したものである。

参考文献

- [1] D. Gooch, A. Wolff, G. Korteum, and R. Brown, "Reimagining the role of citizens in smart city projects," Proceedings of UbiComp/ISWC15 Adjunct, pp.1587–1594, Sept. 2015.
- [2] A.T. Campbell, S.B. Eisenman, N.D. Lane, E. Miluzzo, and R.A. Peterson, "People-centric urban sensing," Proceedings of the 2nd Annual International Workshop on Wireless Internet, Article No. 18, ACM, 2006.
- [3] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M.B. Srivastava, "Participatory sensing," Proceedings of ACM workshop of World-Sensor-Web, pp.117–134, ACM Sensys, Oct. 2006.
- [4] S. Hara, S. Kobayashi, and M. Abe, "Sound collection systems using a crowdsourcing approach to construct sound map based on subjective evaluation," IEEE ICME Workshop on Multimedia Mobile Cloud for Smart City Applications (MMCloudCity-2016), W136, pp.1–6, Seattle, WA, USA, July 2016.
- [5] R.K. Rana, C.T. Chou, S.S. Kanhere, N. Bulusu, and W. Hu, "Ear-Phone: An end-to-end participatory urban noise mapping system," Proceedings of The 9th ACM/IEEE International Conference on Information Processings in Sensor Networks (IPSN 2010), pp.105–116, April 2010.
- [6] E. Kanjo, "NoiseSPY: A real-time mobile phone platform for urban noise monitoring and mapping," Mobile Networks and Applications, vol.15, no.4, pp.562–574, Aug. 2010.
- [7] E. D'Hondt, M.A. Stevens, and A. Jacobs, "Participatory noise mapping works! an evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring," Pervasive and Mobile Computing, vol.9, no.5, pp.681–694, Oct. 2013.
- [8] S. Hara, A. Hatakeyama, S. Kobayashi, and M. Abe, "Sound sensing using smartphones as a crowdsourcing approach," Proceedings of APSIPA Annual Summit and Conference 2017, IEEE, Kuala Lumpur, Malaysia, Dec. 2017. 6 pages.
- [9] S. Kobayashi, M. Abe, and S. Hara, "Prediction of subjective assessments for a noise map using deep neural networks," Proceedings of UbiComp/ISWC 2017 Adjunct, ACM Press, Sept. 2017.
- [10] 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, pp.79–82, 1–Q–7, March 2018.
- [11] A. Mesaros, T. Heittola, A. Diment, B. Elizalde, A. Shah, E. Vincent, B. Raj, and T. Virtanen, "DCASE2017 challenge setup: Tasks, datasets and baseline system," Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop, pp.85–92, Nov. 2017.