GMM-UBM を用いた賑わい識別器の構築

Construction of classifier of bastle using GMM-UBM method

田中 智康

Tomoyasu Tanaka

岡山大学 阿部研究室

Abe Laboratory, Okayama University

概要 本研究ではスマートフォンを用いて集められた環境音から人の賑わいを推定する．賑わいとは人の混雑のことであり，賑わい音を複数人が同時に話すガヤガヤした音と定義する．音収集システムを用いて賑わい音とその他の環境音を収録し，音声認識で一般的な特徴量を用い，GMM-UBM 手法により賑わい識別器を構築・評価した．更に，特徴量に Power を追加した結果，識別性能が向上した．

1 はじめに

人間のまわりを取り巻く音は人間の生活と密接に関わっており，多くの情報が含まれている．音を分析することでユーザに有益な情報を提供することを目的とした研究が盛んである．Hao ら [1] は，睡眠中のいびきなどの音情報を取得し睡眠の質を観測し，ユーザの睡眠パターン提示システムを開発している．これにより，ユーザの睡眠改善を助けることができる．Sun ら [2] の研究では，咳などの 4 種類の音の自動検出システムを開発している．これにより，風邪の流行を把握できる．

本研究では，スマートフォンで収集した環境音から人の賑わいを推定する．賑わいとは人の混雑のことであり，賑わい音を複数人が同時に話すガヤガヤした音と定義する．そこで，賑わい音を識別し，賑わい度を推定して地図上に可視化する手法を検討する．賑わい度の地図上への可視化のイメージを図 1 に示す．楕円が濃い地域ほど賑わっていることを示す．例えば，この地図は地域の住環境を知るための参考になると考えられる．

本報告では，環境音から賑わい度を推定するための賑わい識別器を構築する．賑わい識別器の構築にあたり，環境音を収録し，その環境音データに対して正解となる音の種類のラベルを付与する．そして，ラベルに基づき切り出した環境音を用いて賑わい識別器を構築し，その評価を行う．

2 環境音の収録とラベル付与

2.1 収録データの概要

環境音の収録データは 4 回の収録実験により集められた．収録端末にはタブレット型の Google Nexus 7 を用い，収録者はそれを手に持った状態で収録を行った．1, 2 回目では，岡大から自宅までの往路，岡山駅周辺で収録した [3]．3 回目では，「閑静な住宅街」，「駅前」，「駅に近い商店街」，「駅に遠い商店街」の属性を持つ 4 地域で収録した [4]．4 回目では，成人式（岡山ドーム），センター試験（岡山大学）といったイベント時に収録した．収録により得られた環境音データは 10 秒の wav ファイルが 7499 データ，15 秒の wav ファイルが 194 データの合計 7693 データ (77900 秒) である．収録は量子化ビット数 16 bit，サンプリングレート 32 kHz で行われた．

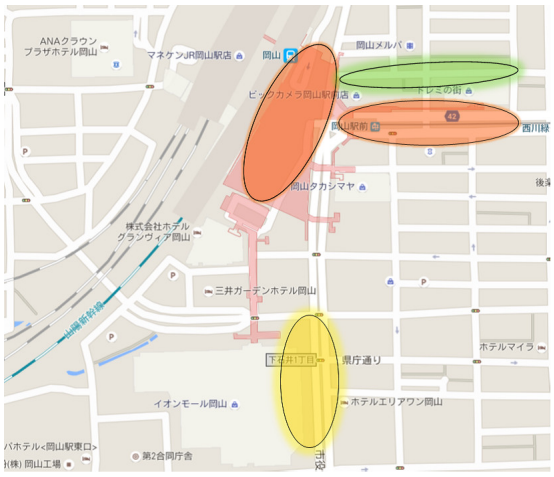

図 1: 賑わい度の可視化のイメージ

2.2 ラベル付与の概要

どの種類の音が賑わい音であると識別される可能性があるのかを調べるため，収録により得られた 7499 データに対してラベル付与を行った．ラベル付与は 2 名で行った．ラベルの種類は，人の声や車の音などの 12 種類の環境音のラベルとした．その後，人のラベルのうち賑わい音であると判断したものは賑わいラベルを付与した．ラベル付与により得られた各クラスのラベル数を表 1 に示す．

3 賑わい識別器の構築

3.1 特徴量の抽出

環境音データの特徴量として，12 次元の MFCC (Mel Frequency Cepstral Coefficient) とその一次差分 ΔMFCC，Power の一次差分 ΔPower の計 25 次元を用いた場合，それらの特徴量の他に Power を追加した 26 次元の場合の 2 パターンで識別器の構築を行う．

表 1: 各クラスのラベル付与数

クラス	ラベル数	クラス	ラベル数
T_{01}（人）	5433	T_{08}（電車）	33
T_{02}（鳥）	3287	T_{09}（サイレン）	130
T_{03}（虫）	186	T_{10}（音響信号機）	6623
T_{04}（車）	4056	T_{11}（動物）	282
T_{05}（風）	1503	T_{12}（音楽）	643
T_{06}（バイク）	559	T_{13}（賑わい）	202
T_{07}（踏切）	34	合計	22971

MFCC は音響特徴量の中でも特によく用いられる特徴量の 1 つで，人の聴覚特性を考慮しながらスペクトルの概形を表現する．なお，分析帯域幅は 0 Hz から 16 kHz とした．MFCC 抽出の際のフレームサイズは 25 msec，フレームシフトは 10 msec，分析窓はハミング窓を用いた．

3.2 GMM-UBM による識別

識別手法は，話者識別で一般的に用いられる GMM-UBM (Gaussian mixture model - Universal background model) [5] を用いた．また，学習モデルには GMM を用いた．GMM は，特徴ベクトルのパターン分布を仮定して学習を行う統計的な手法の一つであり，複数の正規分布の重ねあわせで表される．

GMM-UBM を用いた賑わい識別器の構築法を以下に示す．まず，切り出した環境音を用いて UBM モデル M_{ubm} を GMM として学習する．次に，UBM モデル M_{ubm} を元に賑わいクラス T_{13} のデータへの再学習し，賑わいクラス T_{13} のクラスモデル M_{13} を構築する．最後に，UBM モデル M_{ubm} と賑わいクラスモデル M_{13} の対数尤度比 (Log-likelihood ratio; LLR) を計算する．

ある環境音データ $\boldsymbol{o}$ に対する賑わいクラス T_{13} の LLR は尤度関数 $L(\boldsymbol{o}; M_{13})$ により計算される．

$$LLR(\boldsymbol{o}, T_{13}) = \log L(\boldsymbol{o}; M_{13}) - \log L(\boldsymbol{o}; M_{ubm}) \tag{1}$$

そして，次式によって，賑わいクラス T_{13} の識別器 $C_{13,\theta}(\cdot)$ は与えられた環境音データ $\boldsymbol{o}$ が賑わいクラス T_{13} の環境音である (+1) か否か (−1) を最適な閾値 θ に基づき判定する．

$$C_{13,\theta}(\boldsymbol{o}) = \begin{cases} +1 & \text{if } \ LLR(\boldsymbol{o}, T_{13}) > \theta, \\ -1 & \text{if } \ LLR(\boldsymbol{o}, T_{13}) \leq \theta. \end{cases} \tag{2}$$

学習データには収録したデータに対するラベルに基づいて切り出した 22971 サンプルを用いた．GMM の混合数は 256 として各モデルを作成した．賑わい識別器の学習及び尤度の計算には HTK 3.4.1 [6] を用いた．また，10 分割交差検証により賑わい識別器の学習，性能評価を行う．本報告では，各クラスのサンプル数がいずれのグループにおいても均等になるように 10 分割し，識別器の性能の平均を求めた．

表 2: 分類問題における関係

	実例が正例	実例が負例
予測が正例	True Positive(TP)	False Positive(FP)
予測が負例	False Negative(FN)	True Negative(TN)

4 賑わい識別器の性能評価

本稿では，*F-measure*(F)，適合率 (Pr)，再現率 (Re) を用いて識別性能の評価をおこなう．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を示し，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ものの割合を示す．それぞれの評価指標は表 2 を用いて以下の式により計算される．

$$Pr = \frac{TP}{TP + FP} \tag{3}$$

$$Re = \frac{TP}{TP + FN} \tag{4}$$

$$F = \frac{2Pr \cdot Re}{Pr + Re} \tag{5}$$

25 次元の特徴量（MFCC+ΔMFCC+ΔPower）を用いた場合の賑わい識別器の *F-measure* は 0.584，26 次元の特徴量（MFCC+ΔMFCC+Power+ΔPower）を用いた場合では 0.602 であった．このことから，特徴量に Power を追加した方が識別性能は向上することがわかる．

5 まとめと今後の課題

本報告では，環境音の収録，ラベル付与，GMM-UBM による賑わい推定について述べた．特徴量に Power を追加することで識別性能は向上したが，今後は更なる性能向上に向けた検討が必要となる．

参考文献

[1] T. Hao *et al.*, “iSleep: Unobtrusive Sleep Quality Monitoring using Smartphones,” in ACM SenSys, 2013.

[2] X. Sun *et al.*, “SymDetector: Detecting Sound-Related Respiratory Symptoms Using Smartphones,” UbiComp 2015, pp. 97–108, Sept. 2015.

[3] 原他，“クラウドセンシングにより収集された環境音のシンボル表現を用いた音地図構築手法，”音講論，pp. 1535–1538，Sept. 2014.

[4] 原他，“聴取者の主観評価に基づく音地図作成のための環境音収録，”音講論，pp. 81–82，Mar. 2015.

[5] D.A. Reynolds *et al.*,“Speaker Verification Using Adapted Gaussian Mixture Models,” Digital Signal Processing, vol. 10, pp. 19–41, 2000.

[6] “The HTK Book,” `http://htk.eng.cam.ac.uk/`,