

聴覚障がい者を対象としたスマートグラスを用いた日常生活音の認識

平井 和之 西田 昌史 綱川 隆司 西村 雅史

静岡大学 情報学部

1. はじめに

聴覚障害者は耳が聞こえないため、周囲の状況を把握するのが困難である。聴覚障害者の環境音認知の向上のために視覚情報を利用した環境音学習の有効性について検討されている[1]。環境音の中から緊急の回避や避難を必要とする警告音を識別し、ウェアラブル端末やスマートフォンへの画面表示をすることで、聴覚障害者の日常生活を支援するという研究も行われている[2]。また、環境音を認識し、スマートフォンやタブレットなどの端末に表示させることで環境音を可視化するという研究も行われている[3]。さらに、スマートグラスを用いて、人の音声や環境音を認識し、可視化するシステムの研究が行われている[4]。

従来のスマートグラスを用いた環境音の可視化に関する研究では、環境音の認識方法として SVM (Support Vector Machine) が用いられていた。環境音の認識結果をスマートグラスに可視化するためには、高い認識性能と高速な認識処理が必要となる。そこで、本研究では XGBoost、LightGBM、CatBoost といった勾配ブースティング決定木を用いた環境音の認識手法を提案する。

2. 従来システム

本研究では、環境音の録音および可視化に用いるスマートグラスとして、図 1 の EPSON MOVERIO BP-300 を使用する。



図 1 EPSON MOVERIO BP-300

従来システムでは、スマートグラスに付属しているマイクで人の音声や環境音を拾い、人の音声か環境音かを自動で判別し、人の音声であればその発話内容を音声認識し、環境音であればその音の種類を

識別してスマートグラスに可視化する。また、従来システムでは、サーバー側とクライアント側(スマートグラス)に分けて設計されており、クライアント側では音の録音および可視化を行い、サーバー側では音声か環境音かの判別や音声認識、環境音認識を行う。この一連の流れを以下の図 2 に示す。

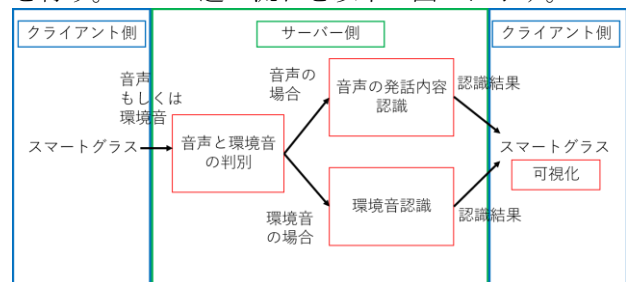


図 2 従来システムの流れ

音声か環境音かを判別する機能と環境音認識の機能には SVM を用いている。音声と環境音はサンプリング周波数 16kHz で収集し、SVM を用いる際の特徴量抽出ではフレーム長 25ms、シフト幅 10ms で抽出した MFCC24 次元を用いている。本研究では、図 2 のうち環境音認識の機能について性能の改善に取り組んだ。

3. 提案手法

本研究では、より高い認識性能と高速な認識処理を実現するために、従来システムで環境音認識に用いられていた SVM を用いた認識手法を、XGBoost、LightGBM、CatBoost といった勾配ブースティング決定木を用いた認識手法に置き換える。

勾配ブースティング決定木とは、今までに学習した決定木によって導出された予測値と実測値との誤差を重点的に学習できるように新たな決定木を構築することでデータの学習を進める方法である。

XGBoost では、勾配ブースティングの訓練過程において、その決定木の成長には「Level-wise」という分岐を行っており、木を階層ごとで同時に分岐させている[5]。LightGBM は、勾配ブースティング決定木に GOSS と EFB を取り入れたものである。GOSS とは、小さい勾配を持つデータに対してランダムサンプリングを行うことでサンプルサイズを減らし、計算量を小さくすることができる。また、EFB はスパースなデータに対して、相互に排他的な特徴量を 1 つにまとめることで情報の損失無く次元数を減ら

Environmental Sound Recognition Using Smart Glass for Hearing Impaired People.

Kazuyuki Hirai, Masafumi Nishida, Takashi Tsunakawa, Masafumi Nishimura

Faculty of Informatics, Shizuoka University

すことができる[6]。CatBoost は、決定木を繰り返し構築する過程で、前回のブースティング段階で毎回新たなデータセットを個別にサンプリングして次の決定木を作るアルゴリズムである[7]。

4. 評価実験

SVM、XGBoost、LightGBM、CatBoost による環境音認識の性能の評価実験を行った。

評価実験に用いる音声データとして、文献[4]で用いられた環境音データ7種類（体温計の音、携帯の着信音、目覚まし時計のアラーム音、レンジの音、ケトルのボタンの音、インターホンの音、ドアが閉まる音）に加え、新たに6種類の環境音（救急車のサイレン、洗濯機の通知音、お風呂の通知音、冷蔵庫の通知音、緊急地震速報の通知音、炊飯器の通知音）を60個ずつ録音した計780個を用いた。

環境音認識の評価方法を示す。学習データに13種類の環境音をそれぞれ50個、計650個使用し、テストデータにそれぞれ10個、計130個使用した場合にできる6パターンで学習とテストを行い、テストデータがどの程度正しく認識されるかという実験を行った。以下の表1に実験結果を示す。認識性能にはF値を用いており、6パターンそれぞれの認識性能の平均であり、F値はイベント単位である。また、RTF(Real Time Fatcor)はテストデータ1個当たりのRTFの平均である。

表1 評価実験の結果

	SVM	XGBoost	LightGBM	CatBoost
認識性能	0.92	0.92	0.96	0.96
RTF	1.287	0.002	0.008	0.003

認識性能はLightGBM、CatBoostが最も高く、RTFはXGBoostが最も小さい結果となった。また、どのパターンにおいても炊飯器の通知音がお風呂の通知音、ケトルと体温計の通知音が地震速報の通知音として誤認識されていた。以下の表2、3にSVMとCatBoostの混同行列を示す。なお、表2、3において列は予測結果を示し、行は正解を示している。

表2 SVMの混同行列

	お風呂	インターホン	ケトル	サイレン	ドア	レンジ	体温計	冷蔵庫	地震速報	携帯	洗濯機	炊飯器	目覚まし
お風呂	60	0	0	0	0	0	0	0	0	0	0	0	0
インターホン	0	59	0	0	0	0	0	0	1	0	0	0	0
ケトル	0	0	48	0	0	0	0	0	12	0	0	0	0
サイレン	0	0	0	60	0	0	0	0	0	0	0	0	0
ドア	0	0	0	0	60	0	0	0	0	0	0	0	0
レンジ	0	0	0	0	0	59	0	0	1	0	0	0	0
体温計	0	0	0	0	0	0	41	0	19	0	0	0	0
冷蔵庫	3	0	0	0	0	0	0	57	0	0	0	0	0
地震速報	0	0	0	0	0	0	0	0	60	0	0	0	0
携帯	0	0	0	0	0	0	0	0	2	58	0	0	0
洗濯機	1	0	0	0	0	0	0	0	0	0	59	0	0
炊飯器	22	0	0	0	0	0	0	0	0	0	0	38	0
目覚まし	0	0	0	0	0	0	0	0	0	0	0	0	60

表3 CatBoostの混同行列

	お風呂	インターホン	ケトル	サイレン	ドア	レンジ	体温計	冷蔵庫	地震速報	携帯	洗濯機	炊飯器	目覚まし
お風呂	60	0	0	0	0	0	0	0	0	0	0	0	0
インターホン	0	59	0	0	0	0	0	0	1	0	0	0	0
ケトル	0	0	52	0	0	0	0	0	8	0	0	0	0
サイレン	0	0	0	60	0	0	0	0	0	0	0	0	0
ドア	0	0	0	0	60	0	0	0	0	0	0	0	0
レンジ	0	0	0	0	0	59	0	0	1	0	0	0	0
体温計	0	0	0	0	0	0	49	0	11	0	0	0	0
冷蔵庫	1	0	0	0	0	0	0	59	0	0	0	0	0
地震速報	0	0	0	0	0	0	0	0	60	0	0	0	0
携帯	0	0	0	0	0	0	0	0	1	59	0	0	0
洗濯機	0	0	0	0	0	0	0	0	0	0	60	0	0
炊飯器	6	0	0	0	0	0	0	1	0	0	0	53	0
目覚まし	0	0	0	0	0	0	0	0	0	0	0	0	60

5. おわりに

本研究では、日常生活音の認識において、より高い認識性能と高速な認識処理のための認識手法として、XGBoost・LightGBM・CatBoostによる手法を提案した。評価実験の結果、上記の3つの手法とSVMのうち、CatBoostが日常生活音の認識に有効であることが分かった。また、現在実装されている可視化システムでは登録されている環境音以外の音が入力された場合、その音を既に登録済みの環境音のいずれかとして認識されてしまう。この問題を解決するため、登録されていない環境音を除外する機能を実装することが今後の課題である。また、音区間の自動検出の実装も行うことでより実用的なシステムの構築を行う。

参考文献

[1]加藤優, 平賀瑠美, 若月大輔, 松原正樹, 寺澤洋子 “聴覚障害者のための視覚情報を利用した環境音学習の基礎的検討” 情報処理学会研究報告 Vol.2017-AAC-4 No.2 pp.1-5 (2017)

[2]白石優旗 “深層学習を用いた警告音認識による危険信号通知システムの検討” 筑波技術大学テクノレポート 24(1), pp.83-84 (2016).

[3]浅井研哉, 綱川隆司, 西田昌史, 西村雅史, “聴覚障害者支援のための環境音可視化システムの開発” 情報処理学会研究報告 Vol.2019-AAC-9, No.5, pp.1-8 (2019).

[4]錦織勇人, 西田昌史, 綱川隆司, 西村雅史 “聴覚障がい者のためのスマートグラスを用いた音声・環境音の可視化システムの構築” 情報処理学会第83回全国大会講演論文集, 12H-01, pp.4_809-4_810 (2021)

[5]Tianqi Chen et al. “XGBoost: A Scalable Tree Boosting System”, Proceedings of the 22Nd ACM SIGKDD (2016).

[6]Guolin Ke et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”, Advances in Neural Information Processing Systems (2017).

[7]Liudmila Prokhorenkova et al. “CatBoost: Unbiased Boosting With Categorical Features”, Advances in Neural Information Processing Systems (2018).