

令和 6 年 5 月 31 日現在

機関番号：14201

研究種目：基盤研究(C) (一般)

研究期間：2019～2023

課題番号：19K02999

研究課題名(和文) 教室内発話ターンテークングに着目した授業活性度の可視化

研究課題名(英文) Visualization of Classroom Activity Levels Based on Turn-taking in Classroom.

研究代表者

市川 治 (Ichikawa, Osamu)

滋賀大学・データサイエンス学系・教授

研究者番号：00821612

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：一般に学校の授業の品質の評価は難しい課題である。良い授業とは何かという問いは評価者によって観点が大きく異なり、また評価も主観的であるからである。そこで授業活性度という単純化された指標を導入し、これを機械学習モデルによって客観的に推定することに取り組んだ。この指標は授業の良し悪しを測るものではなく、教師と生徒の発話のターンテークングが良好であるか、すなわち授業が対話的に行われているかを測るものである。この指標を教室で収録した音声のみによって推定する機械学習モデルを構築した。本手法は授業評価における一つの側面を見るに過ぎないが、対象授業についての客観的な評価を出力できる点がメリットである。

研究成果の学術的意義や社会的意義

教育の現場において、授業評価の方法は大きく2つに分けられる。ひとつは受講生に対してアンケートを行うものであり、もうひとつは他の講師などが授業を参観した上で評価を行うものである。しかし、これらは主観的な評価であり、評価者が変われば変化するという問題があった。本研究では、授業評価の一部の側面に限定することになってしまうが、評価指標を算出する客観的な条件を定め数値化することが重要であるとの考えに立ち、「授業活性度」という定量的な指標を導入した。これを機械学習モデルによって客観的に算出する仕組みを構築した。

研究成果の概要(英文)：In general, evaluating the quality of school teaching is a difficult task. The question of what constitutes good teaching varies widely from evaluator to evaluator, and evaluation is subjective. Therefore, we introduced a simplified metric called "class activity level" and worked to estimate it objectively using a machine learning model. This metric does not measure the quality of the class, but rather the turn-taking of speech between the teacher and the students, i.e., whether the class is interactive. We constructed a machine learning model to estimate this metric using only audio data recorded in the classroom. Although this metric is only one aspect of class evaluation, the advantage of this method is that it can automatically output an objective evaluation of the target class without human intervention.

研究分野：知能情報学

キーワード：授業評価 授業活性度 音響イベント 機械学習

## 1. 研究開始当初の背景

良い授業とは何であるか？その答えは多岐にわたるが、一般的には、生徒が分かる授業が良い授業であるということになるだろう。それでは、教科の内容を減らし、難易度を下げれば良い授業となるだろうか。それも答えは否である。「分かる」という過程は、難易度を下げたり、教師が熱弁を奮ったりするだけでは不十分で、生徒の頭が回転している状態を作ることこそが重要である。

そのため、昨今では授業にアクティブラーニングを取り入れる試みが活発に行われている。しかし、授業の主導権を生徒に委ねるタイプのアクティブラーニングには懐疑的な教師も少なくない。むしろ、従来の授業形式である教師が中心になって話すスタイルを踏襲し、その中で、より多くの「問いかけ発話」を教師が行い、多くの発言を生徒から引き出すインタラクティブな授業を行うのが良いという意見も根強く存在する。インタラクティブな授業を理想的に行うことができれば、生徒の覚醒レベルは高まり、分かるための気づきが多く生まれる。そこで、この状態を活性度の高い授業と呼ぶことにし、それを客観的に計測する必要があると思いついた。

## 2. 研究の目的

本研究の目的は、授業評価における客観的な評価手法を開発することである。従来は、客観的な指標が不足していたために、組織(学校)として、教員の授業品質を管理することが困難であった。本研究で導入する指標は授業評価の一つの側面にすぎないが、今後、他の観点に拡充していくための最初の1歩となる。

## 3. 研究の方法

授業の活性度を測定する有力な手がかりとして、教室内の音声データを利用する。画像データと異なり、教師や生徒のプライバシーを侵害する程度や可能性が小さいというメリットがあり、収録装置に対する心理的な障壁も比較的少ない。

これまでに行われてきた教室内の音声データの利用の試みは、主にグループ学習を対象とし、生徒一人一人にマイクروفオンを装着したうえで、グループ討議の最中に誰がどのような発言をしたかを音声認識装置によって書き起こすアプローチが多かった。生徒の発言を正確にテキストにすることができれば、生徒の気づきを捉え、分かる過程を可視化できる。

しかし、生徒一人一人にマイクروفオンを装着することは実運用では困難である。遠隔位置のマイクروفオンの使用を前提とすれば、生徒の発話音声の品質は劣化するので、音声認識装置によって完全なテキストにすることは難しい。

そこで、本研究では次のようなアプローチをとった。

- ・ 主たる音声収録機器として、教員に装着したピンマイクروفオンを使用する。無指向性のマイクروفオンを使用することで、生徒の発声も収録することができる。ただし、生徒発声の品質は劣化する。
- ・ 従たる音声収録機器として、教室中央付近に設置したマイクروفオンアレイを使用する。生徒の発声をより正確に拾うためである。ただし、これを使わなくても十分な性能が出るのであれば、省略したい設備という位置づけとなる。
- ・ 収録した音声を機械学習モデルに入力して分類を出力させることで、「教師発声」「生徒発声」「板書」「拍手」「笑い」といった教室内の音イベントを検知する。ただし、音声認識による発声のテキスト化は行わない。生徒発話の音声の品質が低いとプライバシーに配慮するためである。
- ・ 検知した音イベントを統計情報に変換し、それを特徴量として授業活性度の指標を出力する機械学習モデルに入力する。入力の特徴量には、教師と生徒の発話交替の回数などのターンテキング情報が含まれる。

## 4. 研究成果

### (1) 収録データ

2020年9月、滋賀大学附属中学校および滋賀大学附属小学校との協力のもと、教育実習生の授業を中心とした授業音声の収録を5日間にわたって実施した。この期間において、中学校で21回、小学校で2回の授業を収録した。

授業音声の収録には、教師に取り付けられたワイヤレスピンマイクروفオンを主たる收音器として使用した。その他に教室の中央付近に配置した2つのマイクروفオンアレイも使用した。収録された全23授業のうち、本研究では中学校の21授業を対象とした。ただしマイクروفオンアレイの音声データを用いた実験については、音声データに欠損のある1授業を除いた20授業を用いて実験を行った。

### (2) 評価者による活性度ラベルの付与

授業音声収録の現地に立ち会った5名と、収録後の音声を聴取した5名により、各授業について

次の基準で評価を行った。

表 1 授業活性化度の評価基準

$\alpha$	教師と生徒が相互に作用しており、授業は活性化していた
$\beta$	教師と生徒のインタラクションはあったが、標準以上に活性化していたとは言えない
$\gamma$	教師と生徒のインタラクションはあったが、少なかった
$\epsilon$	教師がほぼ一方的に話すスタイルの授業であった

この評価分類を 0 点から 3 点に変換し、さらに評価者ごとの標準化を行い、授業ごとの平均をとった数値が、後述の授業活性化度評価のための機械学習モデルの学習に使用する正解ラベルとなる。図 1 にその具体的な数値を示す。授業番号 8 や 13 や 17 のように明らかに評点が低い授業や、授業番号 1 や 2 や 3 のような評点が高い授業があることがわかる。中間的な平均点となっている授業については上下の振れ幅が大きい。人間の評価者による評価はばらつきが大きいですが、明らかに評点が低い授業や高い授業を、本研究の機械学習モデルにより特定できれば、実用的なメリットはあると考える。

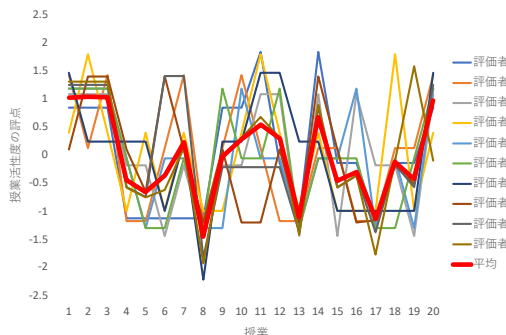


図 1 補間・標準化後の評価点

(3) 授業内音イベント検知のための機械学習モデルの構築

授業内の音イベントは、「教師発声」と「板書」のようにオーバーラップして発生することがあるため、LSTM と LightGBM による 2 段階の機械学習モデルを構成する。図 2 に教師用ピンマイクのみを使用したケースのモデル構成図を示す。1 段階目の LSTM モデルは、時間方向に小分割した音声セグメントの音響特徴量を入力とし、そのセグメントに紐づけられた音イベントの分類を出力する深層学習モデルである。想定される音イベントの種類を表 2 に示す。このモデルは one-hot で学習されるので、そのまま推論に使用した場合には、1 つの音声セグメントあたり 1 つの音イベントしか判定結果として返さない。そこで、このモデルの出力層の値を事後確率の分布 (ベクトル) とみなして、それを複数セグメント結合し、後続の LightGBM モデルに入力する。LightGBM モデルは想定する音イベントごとに学習するので、対象のセグメントを複数の音イベントに判定することが可能となる。ここでは本来連続性を持つコンテキストを判断に含めるために、前後の複数のセグメントを結合してモデルに入力し、真ん中のセグメントを判定対象としている。

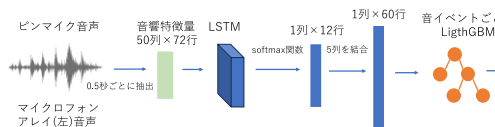


図 2 教師ピンマイクのみを用いた音響イベント検知モデル

表 2 授業内の音イベントの種類

日本語名	英語名
板書	Board Writing
教師発声	Teacher Speaking
男性教師発声	Male Teacher Speaking
女性教師発声	Female Teacher Speaking
教師個別対話	Individual Interaction with Teacher
男性教師個別対話	Individual Interaction with Male Teacher
女性教師個別対話	Individual Interaction with Female Teacher
生徒発声	Student Speaking
グループワーク	Group Work
笑い声	Laughter
拍手	Applause
動画視聴	Video Watching
イベント無し	No Event

(4) 授業活性化度評価のための機械学習モデルの構築

前述の授業内音イベント検知モデルにより、1 つの授業の中で発生するさまざまな音イベントの開始時刻・終了時刻を知ることができる。これを、授業活性化度を推定するモデルへの説明変数 (特徴量) となるように加工する。具体的には、それぞれの音イベントについて、1 つの授業の中で次の統計を算出し、説明変数とする。

- ・ 時間合計
- ・ 回数合計

また、教師と生徒のターンテーキングを表現できるように変数を拡張した (表 3)。

さらに、マイクアレイを使用するケースについては、生徒の発声位置が教室内の特定の場所に偏っていないかを数値化し、授業活性化度推定の説明変数に加える。これは、授業において同じ生徒ばかりが発言をする場合には、授業活性化度は低く評価されるべきという仮定に基づいている。

表 3 ターンテーキング説明変数

拡張した説明変数
生徒の発話時間÷教師の発話時間
教師→生徒のターンテーキング回数
教師→生徒のターンテーキング発生率
生徒→教師のターンテーキング回数
生徒→教師のターンテーキング発生率

教師の発話に対して生徒が反応した (話した・笑った・拍手した) 割合

教師の発話に対して生徒が反応した (話した・笑った・拍手した) 割合

具体的な数値の求め方としては、教室内を 3×3 の 9 カ所の位置ブロックに分割し、マイクロフォンアレイにより求めた生徒発話の発声位置がどのブロックであるかを集計することで、ブロックごとの発話確率を求め、それを発話位置情報量として算出したものである。これらの説明変数を入力とし、人手作成した授業活性度の数値を正解ラベルとして与えて、LightGBM の回帰モデルを学習する。ただし、前述の説明変数はとても種類が多く、そのままモデルの学習に使用すると過学習となる可能性がある。そのため、Lasso による変数選択を使用し、LightGBM モデルに入力する説明変数を十数個に減らしておく。

#### (5) 実験結果と考察

本研究の授業活性度推定モデルでは LightGBM の回帰モデルを用いていることから MSE (平均二乗誤差) による評価を行った。表 4 は交差検証を行った際の MSE の平均値を比較したものである。表中で、入力特徴量：正解ラベルとなっているケースは、説明変数を算出する際の音イベントの情報として、授業内音イベント検知モデルにより予測された結果を使わず、人手により付与された情報を使用したという意味である。当然ながらそのケースの性能は一番良い (MSE が小さい)。

ピンマイクロフォンだけでなく、マイクロフォンアレイを併用したことで、性能は良くなっている。ただし、マイクロフォンアレイの使用によって得られた発話位置情報量については、その効果が認められなかった。マイクロフォンアレイを併用したポジティブな効果は、生徒の発声をより正確に検知できるということから得られている。

ところで、授業活性度推定において実使用を想定すると、「良い」「普通」「悪い」といった分類ができる使いやすい。そこで推定された授業活性度に閾値を設定し、3 段階の分類問題に見立てて結果を整理した。図 3 は、「入力特徴量にピンマイクロフォンとマイクロフォンアレイの授業音声を用いる」かつ「発話位置情報量は説明変数として使用しない」ケースの混同行列である。これは、予測された音イベントラベルを使用する設定の中で最も MSE が小さかったケースとなる。閾値は-0.5 と+0.5 に設けることで 3 段階の分類となっている。混同行列を見ると、実際の授業活性度が 0 のものを 2 と予測することや、2 のものを 0 と予測するといった大きな誤りがないことがわかる。これは、実用性を示唆できる結果であると考えられる。

表 4 授業活性度推定モデルの MSE による性能比較

入力特徴量	正解ラベル		予測ラベル (ピンマイク)		予測ラベル (ピンマイク+アレイ)	
	使用しない	使用する	使用しない	使用する	使用しない	使用する
MSE	0.300	0.441	0.452	0.476	0.354	0.382

		Predicted Class			recall	
		0	1	2		
Correct Class	0	2	2	0	0.50	
	1	3	6	1	0.60	
	2	0	3	3	0.50	
		precision	0.40	0.55	0.75	

図 3 混同行列 (予測ラベル [ピンマイクロフォン+アレイ]、発話位置情報量は使用しない)

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計16件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 高瀬壮彦, 市川治
2. 発表標題 授業内音声イベント検知モデルと授業活性度推定モデルの統合
3. 学会等名 日本音響学会2023年秋季研究発表会
4. 発表年 2023年

1. 発表者名 亀岡 真毅, 市川 治
2. 発表標題 コールセンターオペレーターの感情認識におけるマルチモーダル学習と多言語基盤の効果
3. 学会等名 情報処理学会 第86回全国大会
4. 発表年 2024年

1. 発表者名 川北 海斗, 今井 貴史, 市川 治
2. 発表標題 文意系列を考慮したコールセンターオペレータの発話の文意予測
3. 学会等名 情報処理学会 第22回情報科学技術フォーラム(FIT)
4. 発表年 2024年

1. 発表者名 Osamu Ichikawa, Yuuto Shima, Takahiro Nakayama, Hajime Shirouzu
2. 発表標題 A Method for Estimating the Grouping of Participants in Classroom Group Work Using Only Audio Information
3. 学会等名 2022 IEEE International Conference on Acoustics, Speech and Signal (国際学会)
4. 発表年 2022年

1. 発表者名 Shunya Takano, Osamu Ichikawa
2. 発表標題 Automatic scoring of short answers using justification cues estimated by BERT
3. 学会等名 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022 (国際学会))
4. 発表年 2022年

1. 発表者名 中野 魁人, 高瀬 壮彦, 市川 治
2. 発表標題 授業内音声イベント検知による授業活性度の推定
3. 学会等名 日本音響学会 2022年秋季研究発表会
4. 発表年 2022年

1. 発表者名 高瀬 壮彦, 市川 治
2. 発表標題 教員ピンマイクによる授業内音声イベントの可視化
3. 学会等名 日本音響学会 2022年春季研究発表会
4. 発表年 2022年

1. 発表者名 徳永 一輝, 中山 隆弘, 白水 始, 市川 治
2. 発表標題 対話の構造の可視化のためのBERTを用いた先行発話への接続確率の推定モデル
3. 学会等名 情報処理学会全国大会
4. 発表年 2022年

1. 発表者名 島 悠人, 中山 隆弘, 白水 始, 市川 治
2. 発表標題 授業グループワークの参加者を音声情報のみを用いて自動グルーピングする方法
3. 学会等名 日本音響学会秋季研究発表会
4. 発表年 2021年

1. 発表者名 高野 峻矢, 市川 治
2. 発表標題 BERTによる国語記述式答案の自動採点における採点アノテーションの重要性
3. 学会等名 第20回情報科学技術フォーラム
4. 発表年 2021年

1. 発表者名 Osamu Ichikawa, Kaito Nakano, Takahiro Nakayama, Hajime Shirouzu
2. 発表標題 Multi-Channel VAD for Transcription of Group Discussion
3. 学会等名 Interspeech 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 藤岡和紘, 小松 美幸, 市川 治
2. 発表標題 Gesture-to-talkによるサインエージ端末用の音声区間検出の改善
3. 学会等名 情報処理学会 第83回全国大会
4. 発表年 2021年

1. 発表者名 柿木 博行, 白水 始, 中山 隆弘, 市川 治
2. 発表標題 グループワーク音声可視化のための疑問符( ?マーク )の自動挿入
3. 学会等名 日本音響学会秋季研究発表会, 2020
4. 発表年 2020年

1. 発表者名 花房 万由子, 市川 治
2. 発表標題 TSP信号とマルチマイクロフォンによる打音検査診断の精度向上
3. 学会等名 日本音響学会秋季研究発表会, 2020
4. 発表年 2020年

1. 発表者名 高山 和明, 市川 治
2. 発表標題 音声感情認識の認識精度に対する発話末尾の音響特徴量の影響
3. 学会等名 日本音響学会秋季研究発表会, 2020
4. 発表年 2020年

1. 発表者名 水口 綾乃, 市川 治
2. 発表標題 分散表現とLSTMを使用した滋賀大学オープンキャンパス来訪者向け質問応答システムの構築
3. 学会等名 情報処理学会 第82回全国大会
4. 発表年 2020年



〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 発話者グループ推定装置	発明者 市川 治	権利者 滋賀大学・内田 洋行
産業財産権の種類、番号 特許、特願2021-132459	出願年 2021年	国内・外国の別 国内

〔取得〕 計1件

産業財産権の名称 マルチチャンネル発話区間推定装置	発明者 市川 治	権利者 滋賀大学
産業財産権の種類、番号 特許、特願2020-004771	取得年 2021年	国内・外国の別 国内

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	伊達 平和 (Date Heiwa) (70772812)	滋賀大学・データサイエンス学系・准教授  (14201)	
研究分担者	右田 正夫 (Migita Masao) (70335157)	滋賀大学・教育学系・教授  (14201)	
研究分担者	田中 宏子 (Tanaka Hiroko) (00324559)	滋賀大学・教育学系・教授  (14201)	
研究分担者	糸乗 前 (Itonori Saki) (90324558)	滋賀大学・教育学系・教授  (14201)	
研究分担者	辻 延浩 (Tsuji Nobuhiro) (00378431)	滋賀大学・教育学系・教授  (14201)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	磯西 和夫  (Isonishi Kazuo)  (50223061)	滋賀大学・教育学系・教授     (14201)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関