

CO2センサーを利用した教室状態解析システムの開発

黒羽 晟¹ 高田 晃希¹ 栗原 諒¹ 爰川 知宏^{1,†1}

概要: 本研究は東京国際工科専門職大学の教室に設置された M5StickC によるセンシングから教室状態遷移の解析を行い、教室状態解析システムを開発した。解析では教室からセンシングした CO2 濃度データなどを入力とし、教室使用確率を出力とする機械学習モデルを構築した。本研究の目標は構築した機械学習モデルを組み込んだ教室状態解析システムにより、学生の教室移動や時間割管理を支援することである。

1. はじめに

現在、会社や学校の予定はスケジュール表、時間割などの静的な情報に依存している。これらは事前に定義されており、突発的な例外により時間割などに沿った予定通りにならないことは少なくない。例外による予定変更気づかずに会議室や教室の移動を行ってしまうと遅刻及び欠席評価に繋がる恐れがある。このような課題を解決するために人感センサーの複数設置、カメラを設置して深層学習モデルによる推論を行う手法などが考えられるが、費用がかかる。

そこで本研究では、教室を複数人が利用すると CO2 濃度が増加するという仮説の元、カメラなどに比べて安価な温湿度 CO2 センサと機械学習モデルを用いて教室状態解析を行い、学習したモデルを活用して、東京国際工科専門職大学（以下、本学と呼ぶ）の学生が利用可能な教室状態解析システムを開発した。教室状態解析システムは時間割の情報と時間割に基づいた受講予定の教室の使用確率を動的に確認する機能要件を満たすもので、Web ブラウザで使うことができる。

本研究の目的は、時間割通りでない例外的な授業教室変更の課題に対して学生が対応可能なシステムの開発と当該課題を解決する上でのインフラ費用削減とする。

2. 既往研究

部屋の状況、特に居住状況を分析するために、様々な研究がセンサーの利用を模索している。Szczurek ら [1] と [2] はそれぞれ、人の存在を検知するためのガスセンサーと環境センサーの可能性を強調している。Lam ら [3] はこれを一歩進め、空間の居住者数を検出するアルゴリズムを開発

している。R ら [4] はこの研究を発展させ、非侵入型センサーと機械学習技術を用いて、部屋の占有率を高い精度で推定することに成功している。これらの研究は総体的に、ビル管理とエネルギー効率を最適化するためのセンサーベースの分析の可能性を示している。

3. 教室状態解析システム

3.1 概要

教室状態解析システムは、図 1 のように本学の教室が時間割通りに使用されているかどうかを 3 段階の色で教室番号を表示する。



図 1 ダッシュボード画面

時間割通りに教室が使用されているかどうかを示す指標を信頼度として定義する。信頼度 85%以上で青色、信頼度 85%未満かつ 40%以上で黄色、信頼度 40%未満で赤色を表示する。また、色で表示するだけでなく色覚異常のユーザーのために、カーソルを合わせると信頼度が表示される画面に設計した図 2。

信頼度について説明する。温湿度 CO2 センサーにより計測したデータを入力として機械学習モデルが 0.0 - 1.0 の範囲で出力した教室使用確率の予測値を \hat{y} 、真の値を y とし、予測値と真の値の差分の絶対値 $|\hat{y} - y|$ を誤差 e 、最大

¹ 東京国際工科専門職大学

^{†1} 現在、東京国際工科専門職大学



図 2 ダッシュボード画面

許容誤差を r として信頼度 C の計算式 (1) を示す.

$$C = (1 - e/r) \times 100 \quad (1)$$

時間割を元に作成した教室データを真の値としてこの信頼度を算出した.

3.2 温湿度 CO2 センサーによる計測

M5StickC に温湿度 CO2 センサーを接続したものを本学の教室に設置されており [5], 学内 Wi-Fi から InfluxDB の API を利用して機械学習モデルの訓練データ用に 12 月 6 日から 12 月 14 日まで, 最終モデル検証用に 12 月 18 日~12 月 20 日までの計測時刻, 温度, 湿度, CO2 濃度のデータを csv 形式で取得した. この取得した csv データは 30 秒間隔の計測データになっている. 対象とした教室は 4 教室である.

3.3 教室データ解析

まず最初に, 学習データの特徴を把握するために EDA(Explanatory Data Analysis) を行った. 教室状態確認システムで利用する機械学習モデルは教師あり回帰モデルを採用し, 学習データの目的変数である教室使用確率は次のように作成した. 時間割に基づいて, その時間教室が使用されていれば 1.0, 使用されていなければ 0.0, 教室が使用された後の休み時間であれば 0.7, 使用された後のお昼休みであれば 0.3 とした. 休み時間は次に授業があれば殆どの学生が移動せず, 帰宅の際は休み時間の 15 以内に一齐に帰宅することは少ないため 0.7 とした. お昼休みは殆どの学生が昼食に向かい, 教室が空き状態に近くなるため 0.3 とした.

また, 土日と 1 限から 7 限以外のデータは削除し, 機械学習に入力する学習データとしてテーブルデータにした.

機械学習に入力する訓練データは 29115 レコード, 5 カラムであり最終的な検証に用いるデータは 14894 レコード, 5 カラムでカラムの詳細は以下の (表 1) の通りである.

表 1 カラムの詳細

説明変数	説明	型
co2	CO2 濃度	float 型
humidity	湿度	float 型
temperature	温度	float 型
y	教室使用確率 (目的変数)	float 型

3.4 機械学習モデルの構築

ハイパーパラメータチューニングを行うモデルを数ある回帰モデルの中から選定するため, 自動機械学習ライブラリ PyCaret を用いて 25 種類のモデルを 5 分割交差検証により学習し, 6 つの指標でモデルの性能を評価した. Extra Trees が最も精度が良く, Extra Trees ハイパーパラメータチューニングを行うモデルとして選定した.

機械学習ライブラリ scikit-learn の Extra Trees 回帰モデルに対し, ハイパーパラメータの自動最適化フレームワークである Optuna でベイズ最適化によるハイパーパラメータチューニングを行った. さらに, 精度に応じてハイパーパラメータだけでなく EDA の結果から前処理及び特徴量エンジニアリングを施した.

3.5 モデルの API 化

採用した PyCaret で学習したモデルを pickle 形式で保存し, API 化した. CO2, 温度, 湿度, 座席数を json 形式で HTTP の POST リクエストすると, 教室使用確率の予測値がレスポンスされるようにした.

4. 結果と考察

4.1 EDA 結果と考察

EDA では, 相関分析, 箱ひげ図による外れ値分析, 曜日や時間と co2 の関係を視覚化した.

相関分析の手法として相関行列を算出し, ヒートマップにして視覚化した図 3. 相関行列ヒートマップを見ると, 相関係数が最も大きいのは CO2 と目的変数であることが読み取れ, この結果は, 時間割通りに教室を複数人が使用することで対象の教室の CO2 濃度が増加した結果であると考えられる.

箱ひげ図による外れ値分析は四分位数の情報から, 第一四分位数 - 1.5 × 四分位範囲 をひげの下限, 第三四分位数 + 1.5 × 四分位範囲 をひげの上限とし, ひげの上端より大きい値とひげの下端より小さい値を外れ値として扱った図 4.

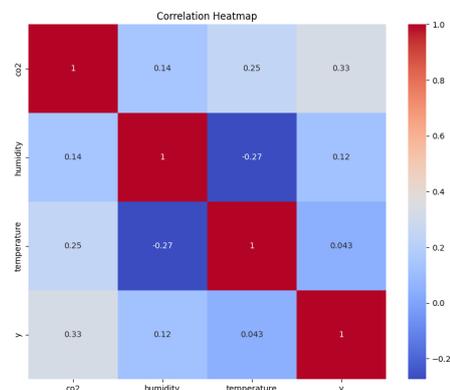


図 3 相関行列ヒートマップ

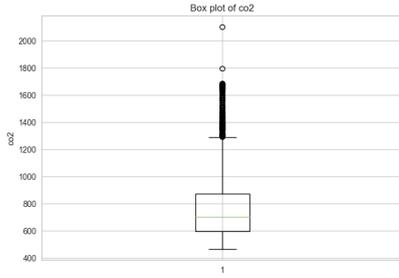


図 4 箱ひげ図

4.2 Pycaret によるモデル性能評価結果と考察

自動機械学習ライブラリ PyCaret を用いて 25 種類のモデルを 5 分割交差検証により学習し、6 つの指標でモデルの性能を評価した結果、モデルの性能が上から良い順に 3 つ並べたものを (表 2) に示す。6 つの指標は MAE, MSE, RMSE, R^2 , RMSLE, MAPE である。RMSE は MSE の平方根であり、MAPE は真の値に 0 を含む場合を定義していないため (表 2) では MSE と MAPE は割愛した。

式 (2) に MAE, 式 (3) に RMSE, 式 (4) に R^2 , 式 (5) に RMSLE の定義式を示す。ここで、式中の n はレコード数、 \hat{y}_i は教室使用確率の予測値、 y_i は時間割に基づいた真の値を表す、

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2} \quad (5)$$

表 2 性能上位 3

	MAE	RMSE	R^2	RMSLE
Extra Trees	0.0109	0.0575	0.9824	0.0392
Random Forest	0.0153	0.0669	0.9762	0.0458
Decision Tree	0.0114	0.0873	0.9592	0.0595

Extra Trees が 6 つの指標で最も精度が良かった。

4.3 ベイズ最適化によるハイパーパラメータチューニング結果と考察

4 の特徴量に対して標準化を行い、ベイズ最適化によるハイパーパラメータチューニングを行なった結果、検証データに対する予測精度は RMSE が 0.48, 決定係数が 0.14 であった。訓練データの EDA から導き出された外れ値を除外し、曜日と時限を特徴量として追加し、ワンホットエンコーディングしたテーブルデータを入力に変更したところ

予測精度は RMSE が 0.35, 決定係数が 0.38 となり精度向上が見られた。

5. おわりに

温湿度 CO2 センサーからエクストラツリーモデルで教室の使用確率を予測し、そこから時間割の信頼度を算出し表示するシステムを開発した。

しかし、開発側だけで完璧な精度の学習データを作り続けるのは非常に手間がかかるため、ユーザーからのフィードバックを受けられるようにし、信頼性の即時反映および高精度のデータを確保するとともに、リアルタイムで得られたフィードバックを元に fine-tuning することにより、再学習のコストを下げるという機能を実装するつもりである。

謝辞 本開発の遂行におきまして、東京国際工科大学工学部情報工学科 愛川知宏教授よりご提供いただいた、教室内温湿度 CO2 データ及び API は必要不可欠でありました。制約のあるネットワーク環境下にも関わらずこれらを可能とした高度な知識・技術に敬意を表すと共に、深く感謝致します。

参考文献

- [1] Andrzej Szczurek. and Monika Maciejewska. and Tomasz Pietrucha.: *Occupancy Detection using Gas Sensors*, International Conference on Sensor Networks (2017).
- [2] Mohammadhossein Toutiaee.: *Occupancy Detection in Room Using Sensor Data*, ArXiv (2021).
- [3] Khee Poh Lam and Michael Höynck and Bing Dong and Burton Andrews and Yun-Shang Chiou and Rui Zhang and Diego S. Benítez and Joon-Ho Choi: *Occupancy detection through an extensive environmental sensor network in an open-plan office building*, Eleventh International IBPSA Conference (2009).
- [4] Deepa R. and Kousshik Raj. and N. Balaji. Duragadevi K.: *Machine Learning based Estimation of Room Occupancy Using Non-Intrusive Sensors*, International Conference on Communication, Computing and Internet of Things (IC3IoT) (2022).
- [5] T. Kokogawa.: *Trial of Building a Resilient Face-to-Face Classroom based on CO2-based Risk Awareness*, ITDRR (2021)