

複数センサデータの自動可視化に向けたユーザ操作機械学習の予備評価

○藤野慶汰[†] 松木萌[†] 井上創造[†] 柴田智広[†]
[†]九州工業大学

Preliminary Evaluation of User-operation Learning for Automatic Visualization of Multiple Sensor Data

Keita Fujino, Moe Matsuki, Sozo Inoue, and Tomohiro Shibata
[†] Kyushu Institute of Technology

Abstract: The purpose of this study is to automatically visualize multiple sensor data according to the user. Therefore, the user operated to perform zooming and sensor selection on visualized data, and we analyze where the user is examining on. The operation was estimated by using random forest with features from the data as the explanatory variables and the user operations as the objective variable, and was analyzed using important factors. As a result, it was found that features such as “Number of Samples” of data and “Maximum Value” and “Variance” on the right side of visualized data were high importance.

1. はじめに

近年、スマートフォンの普及やセンサデバイスの小型化・低価格化が進み、これらを利用した研究が様々行われている。例えば、照度センサや電力センサを利用し、家庭内での行動をセンシングする研究[1]や位置センサと消費電力センサを利用した行動センシング研究[2]などが行われている。これらで得られた膨大なセンサデータの活用方法として、第一には可視化を行う方法が考えられる。その際、ユーザの職業や研究分野によってもデータの着目点が異なるため、ユーザに適したデータの可視化を行う必要がある。そこで本研究では、このような複数センサデータを閲覧時のユーザの行動データを収集し、教師あり学習アルゴリズムを適用することにより、自動的に適切な可視化を実現することを最終的な目的である。本稿では、そのためのコンセプトを提案し、予備的に行った実験とその結果を示す。

2. 提案コンセプト

本研究における自動可視化は、(1)可視化アルゴリズムの準備、(2)操作行動データ収集、(3)教師あり学習アルゴリズムの適用、により実現する。(1)は、まず各センサ毎に予め必要と考えられる可視化アルゴリズムを準備し、続いて複数のセンサデータの中から一部を選択する、もしくは複数のデータの関係性を分析するなどの多様なアルゴリズムを準備することを指す。(2)は、(1)で準備した可視化アルゴリズムをユーザが用いる操作行動データを収集することを指す。(3)は、(2)で得た操作行動データを教師データとし、その際に用いた可視化特徴量を入力として、教師あり学習アルゴリズムを適用することを指す。これにより、未知のセンサデータに

対してユーザの操作行動を予測し、自動的に適切な可視化を実現することができる。本稿では具体的に、2種類のセンサデータを可視化するケースを用いて予備的な実験を行いその結果を示す。可視化するデータは、我々が現在介護施設で行っている行動分析実験の際に得られている照度センサと温度センサのデータとする。可能な操作行動としては、照度センサ選択、温度センサ選択、左ズーム、右ズームという、計4種類とする。より詳細な実験内容については後述するが、本予備実験では、4種類すべての操作が実行できる場合、照度センサまたは温度センサのみが表示されて左右ズームの操作のみができる場合、センサ選択の操作のみでできる場合、に分類してそれぞれの場合の特徴量を計算する。この特徴量を説明変数に、ユーザが行った操作を目的変数としてランダムフォレスト法で操作を推定する。更に、その時に得られた重要度を用いて特徴量の評価を行う。

3. 実験

3.1 介護施設での実験

実験内容は、スマートフォンの照度センサ及びセンサタグの赤外線温度センサを利用してリアルタイムでセンサ情報を収集する。この収集したデータをもとに人の行動分析などを行うことが目的である。センサタグは Texas Instruments 社の SimpleLink SensorTag CC2650STK を利用し、利用者個人の部屋のベッド横及び介護職員の胸の部分に設置・装着をする。スマートフォンは Android OS が搭載された端末を利用し、センサタグを設置している部屋及び共用スペースに設置する。予備実験ではここで得られた温度センサと照度センサ合わせて約 150 日分のセンサデータを利用した。

3.2 実験システム

実験を行うにあたり、図1のようなWebシステムを開発した。使用言語はJavaScript、使用ブラウザはGoogle Chromeとした。画面の構成は画面上部に温度センサを、画面下部には照度センサを可視化したものを表示する。さらにその下部には、「温度センサ」と「照度センサ」のボタンが配置されており、表示するセンサを選択することができる。その横には「左ズーム」と「右ズーム」ボタンがあり、左右にズームすることが可能となっている。具体的な実験方法については次項で述べる。

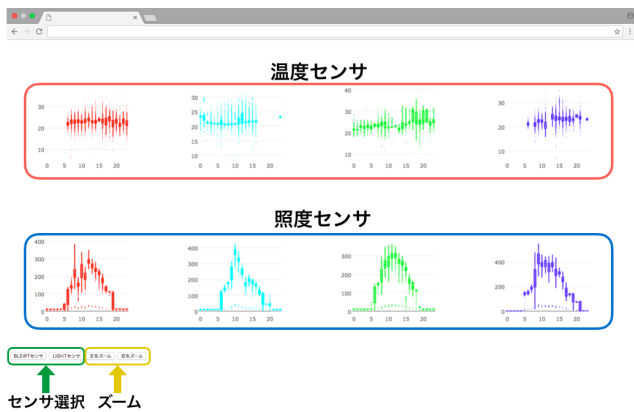


図1 使用する実験システム

3.3 実験方法

可視化されたセンサデータに対してどのような操作をしたかを収集するための方法を図2に示す。可視化するセンサデータは本研究室が現在介護施設で行っている行動分析実験の際に得られている照度センサと温度センサの2種類とした。1日ごとにサンプル数の違う2種類のセンサデータ合わせて約150日分のデータの中からランダムに1日毎に取り出して計4日分を可視化し、被験者にズームやセンサ選択の操作をしてもらった。データを可視化するためのグラフの種類は箱ひげ図とし、図3と図4に示す。箱ひげ図は、最大値や最小値などの基本的な統計量がわかる、データのばらつき具合がわかるなどといった特徴がある。今回は1時間ごとにデータのばらつき具合を見るために箱ひげ図を用いて可視化を行った。可視化されたグラフの操作の種類は、「右側をズーム」「左側をズーム」「温度センサを選択」「照度センサを選択」の計4つの簡単な操作とした。これは、特徴量の次元を抑えることができるためにこれら4種類の簡単な操作とした。特徴量の次元を抑えることにより計算量が減り、学習がしやすくなる。最終的に1種類1日分のセンサデータとなるように被験者にはズー

ム2回とセンサ選択1回の操作を行ってもらった。その際に、データの「サンプル数 (Sample)」「データ日数 (Day)」「最大値 (Max)」「最小値 (Min)」「平均 (Mean)」「中央値 (Median)」「第一四分位点 (1stQu)」「第三四分位点 (3rdQu)」「分散 (Variance)」「標準偏差 (Sd)」の10種類の特徴量を取得した。この作業を1サイクルとして、23歳男性の理系大学院生3人の被験者に1人あたり50サイクル計150サイクル分のデータを収集した。

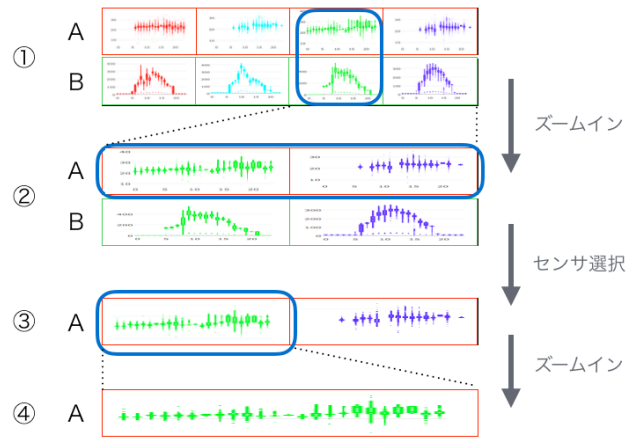


図2 可視化されたデータの操作方法の例

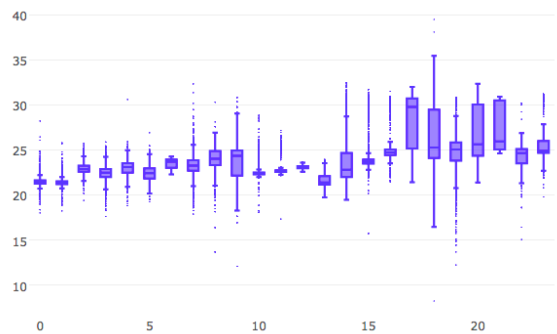


図3 温度センサの可視化の例

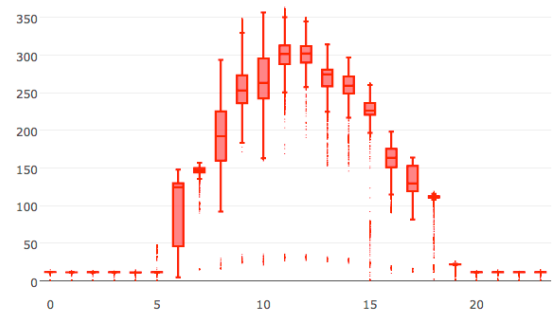


図4 照度センサの可視化の例

3.4 実験結果

被験者ごとに50サイクル150操作の内訳を表1に示す。被験者AとBの場合は、右側をズームする回数がわずかに多くなった。また、全ての被験者がセンサ選択の操作においてほとんどが照度を選択する結果となった。この結果から得られた特徴量とユーザ操作からランダムフォレスト法で評価を行った。

表1 被験者ごとの操作の内訳

操作 \ 被験者	A	B	C
右ズーム	51回	59回	50回
左ズーム	49回	41回	50回
照度選択	50回	46回	37回
温度選択	0回	4回	13回

4. 評価

4.1 評価方法

評価方法としてR言語のrandomForestパッケージを利用する。被験者が行った操作を目的変数に、それ以外の特徴量を説明変数として操作を推定する。さらに、ランダムフォレスト法によって得られる特徴量の重要度から分析を行う。今回は収集されたデータセットを以下の4つに分類してそれぞれ評価を行った。

1. すべてのデータが表示されている場合
2. 照度センサのみ表示されている場合
3. 温度センサのみ表示されている場合
4. センサ選択の操作のみ残っている場合

1に関しては4つすべての操作ができる状態、2と3に関してはズームのみができる状態、4に関してはセンサ選択のみができる状態とする。また、パラメータである各決定木を作るときに使う特徴量の数はそれぞれ6, 4, 8, 4とし、決定木の数は50とした。

4.2 評価結果

4.1節で述べた各状態における混同行列を表2-5に示す。各行は実際の操作を示し、各列は予測された操作を示している。また、各状態における特徴量の重要度を図5-8に示しており、縦軸は重要度、横軸は特徴量を表しており、Lightは照度センサ、IRTは温度センサ、RightとLeftはそれぞれ可視化されているデータの右側と左側を表している。4.1節で述べた1の場合の混同行列と特徴量の重要度を表2と図5に示す。この場合は4つすべての操作ができるため、チャンスレベル25%であるが

分類した結果の正解率は43.8%であった。特徴量の重要度は「サンプル数」が最も重要度が高く、その次に表示データ右側の特徴量の重要度が高くなっている。4.1節で述べた2の場合の混同行列と特徴量の重要度を表3と図6に示す。照度センサのみが表示されている場合は左右どちらかのズーム2択なので、チャンスレベル50%であるが分類した結果の正解率は76.9%であった。特徴量の重要度が高いものは、表示されているデータ右側の特徴量となっており、反対に重要度が低いものは、データの「最小値」と「データ日数」であった。4.1節の3の場合の混同行列と特徴量の重要度を表4と図7に示す。この場合も4.1節の2の場合と同じく左右どちらかのズーム2択なので、チャンスレベル50%であるが分類した結果の正解率は77.3%となった。特徴量の重要度が高いものは、こちらも表示されているデータ右側の特徴量の重要度が高くなっており、重要度の低いものも同じようにデータの「最小値」と「データ日数」となっている。4.1節の4の場合の混同行列と特徴量の重要度を表5と図8に示す。この場合は照度センサか温度センサのどちらかを選ぶ2択なので、チャンスレベル50%であるが分類した結果の正解率は88.2%であった。また、重要度の高い特徴量は温度センサのデータ特徴量となっており、「サンプル数」も比較的重要度が高くなっている。

表2 すべての操作が可能な場合の混同行列

真 \ 推定値	右ズーム	左ズーム	照度選択	温度選択
右ズーム	43	15	30	0
左ズーム	19	22	28	0
照度選択	23	17	47	1
温度選択	5	1	5	0

表3 ズーム操作のみ可能な場合の混同行列 (照度)

真 \ 推定値	右ズーム	左ズーム
右ズーム	48	14
左ズーム	14	45

表4 ズーム操作のみ可能な場合の混同行列 (温度)

真 \ 推定値	右ズーム	左ズーム
右ズーム	7	3
左ズーム	2	10

表 5 センサ選択のみ可能な場合の混同行列

真 \ 推定値	照度選択	温度選択
照度選択	45	1
温度選択	5	0

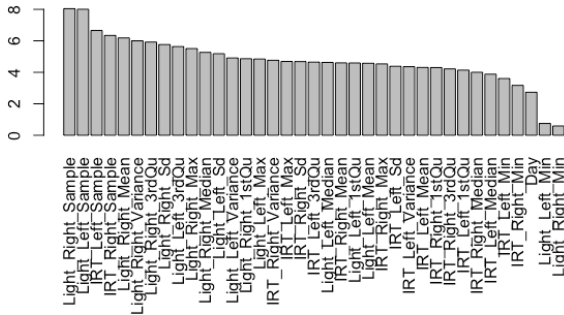


図 5 すべての操作が操作可能場合の重要度

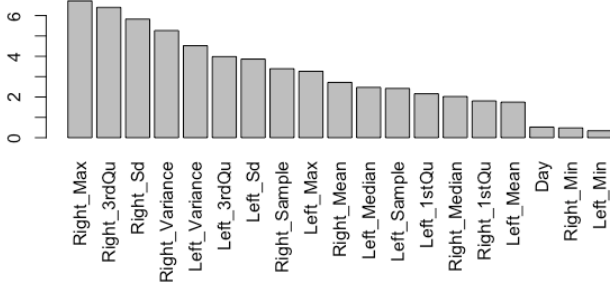


図 6 ズーム操作のみ可能な場合の重要度 (照度)

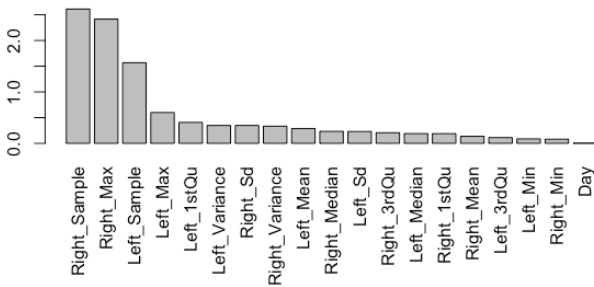


図 7 ズーム操作のみ可能な場合の重要度 (温度)

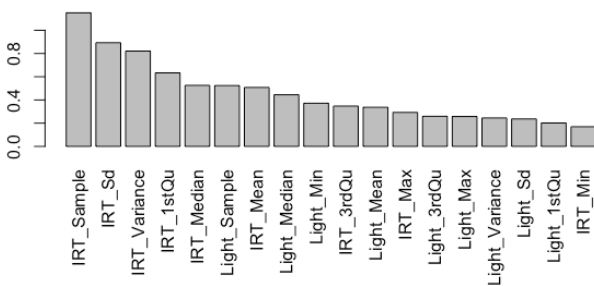


図 8 センサ選択のみが可能な場合の重要度

5. 考察

今回、センサデータに対するユーザの操作を目的変数に、センサデータの基本的な統計量を説明変数としてランダムフォレスト法を利用し、評価を行った。その結果、4.1節の1の状態では、各データの「サンプル数」を特徴量としたものが高い重要度となった。これは、操作を行う際にデータのサンプル数が多いほどその部分のデータを注目していると推測できる。4.1節の2の状態では、データ右側の「最大値」や「第3四分位数」、「標準偏差」などの特徴量が、4.1節の3の状態では、データ右側の「サンプル数」や「最大値」などの特徴量が高い重要度となっている。これは、予備実験でもデータの左側よりも右側にズームする回数が増えているため、右側のデータに注目してズームしていることが推測できる。またこれら3つの状態では「データ日数」「データの最小値」の特徴量は重要度がかなり低いことが示された。4.1節の4の状態ではほとんど照度センサが選択されていたが、温度センサが選択されたときは照度センサよりも「サンプル数」が多い時であった。図8でも温度センサの「サンプル数」が重要度の高いものとなった。

6. まとめ

本研究では、複数センサデータを可視化し、ユーザにズームとセンサ選択の操作を実際に行ってもらった。その結果からランダムフォレスト法を利用して分析したところ、ユーザがズームやセンサ選択を選ぶ基準としてサンプル数が最も重要であることが分かった。さらに、左右どちらかを選ぶ二者択一の場合は、右側にあるセンサデータのほうがわずかであるが選ばれやすく、特徴量も右側のデータのほうが重要という結果になった。今回は可視化するためのグラフの種類を箱ひげ図を利用したが、今後は、他の種類でも実験する必要がある。さらにセンサの種類を増やす、ズーム機能の複雑化、サンプル数がかかなり多い場合などでも分析をしていき、対象者に合わせた自動可視化を目指す。

参考文献

- [1] 井上 創造, 潘 新程, 花沢 明俊, 家庭内行動センシングにおける機械学習データの家庭間転移について, 情報処理学会ユビキタスコンピューティングシステム(UBI)研究報告, pp. 8 pages, 2015/11/29, Beppu.
- [2] 上田健揮, 玉井森彦, 荒川豊, 諏訪博彦, ユーザ位置情報と家電消費電力に基づいた. 宅内生活行動認識システム, 情報処理学会論文誌, Vol.57, No.2, p.1-10 (Feb. 2016).