

センサデータの可視化におけるユーザ着目点の機械学習

藤野 慶汰¹ 井上 創造¹ 柴田 智広¹

概要：膨大なセンサデータを観察者の専門性や目的に応じて適切に可視化することで、観察者の作業効率化を図ることができると期待される。そこで我々は、センサデータ閲覧時のユーザの着目点情報を収集し、表示されているセンサデータの特徴量を説明変数として、教師あり機械学習アルゴリズムの一つであるランダムフォレスト法を適用することにより、各データについてどの特徴量が着目されるかを推定する手法を提案している。本稿では、我々が介護施設での実験で得た、照度と加速度データを用いて行った学習実験および評価実験について述べる。評価実験の結果、照度センサデータに対しては、可視化手法によらずユーザはデータの変化が大きい部分に着目するため、高精度で着目点推定が可能であることが分かった。一方、加速度センサデータに対しては、ユーザによっても、可視化手法によっても着目点異なることが分かった。今後、加速度センサに対して良い着目点推定を行う可視化手法や特徴量を探索する必要がある。

Machine Learning of User Attentions in Sensor Data Visualization

KEITA FUJINO¹ SOZO INOUE¹ TOMOHIRO SHIBATA¹

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の研究 [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、第一には可視化が考えられる。その際、観察者の専門性や目的によってもデータの着目点異なることが考えられるため、単なる教師なし学習ではなく、観察者の視点を教師あり機械学習により考慮に入れた可視化が期待される。この膨大なセンサデータを観察者の専門性や目的に応じて適切に可視化することによって作業の効率化を図ることができる。そこで本稿では、センサデータ閲覧時のユーザの着目点を収集し、教師あり機械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手法を提案する。その結果、データの変化が明瞭のセンサデータは高い精度で着目点を推定することができた。しかし、データが常に変動し続けるようなセンサデータの場合には推定精度が少し低くなったが、両者ともチャンスレ

ベルよりも高くなった。

本論文での貢献は以下のとおりである。

- 膨大なセンサデータからユーザが着目すると考えられる部分のみを可視化できる手法を提案し、その精度を示した。この手法はユーザの視点を教師あり学習するという、教師なし学習による可視化とは異なるアプローチを取っている。
- ユーザの着目にどのセンサ特徴量が寄与するかを、アンサンブル学習による重要度から抽出した。
- 本手法は画像からの着目点の学習とは異なり、あるセンサデータの異なる可視化手法にも適用できる方法である。そこで、学習した推定モデルを、別の可視化手法にも適用できるかどうかを検証した。

2. 関連研究

データの可視化を行う場合に、どの可視化が適切か推薦するシステムが提案されている [3]。しかし、センサデータなどの時系列データの場合、一般的に折れ線グラフを用いられることが多い。可視化手法に限られるためこのシステムを利用することはあまり意味がない。この時系列データに特化したシステムが提案されている [4]。ユーザが可視化された時系列データの着目点をズームすることでそ

¹ 九州工業大学大学院

の部分のみを可視化し、さらにどの部分をズームしたのかがツリー上に表現される。この場合、データの量が膨大になると何が描画されているのか把握できずにズームすることが困難となる。このような問題を解消するために、3次元空間上で時系列データを可視化する手法が提案されている [5]。この手法を利用することで膨大なセンサデータでも可視化することが可能となる。しかし、3次元空間上で表現すると2次元平面上の場合と比較すると観察することが困難になる。そこで本稿では膨大なセンサデータを自動的に適切な部分のみを可視化する手法を提案する。

3. 提案手法

提案手法では、センサデータとその可視化結果、さらにその可視化結果を見た人が着目した結果のデータが収集されていることを前提とする。今回は着目した結果のデータとして、可視化されたセンサデータに対してユーザが着目した箇所に赤枠を付けた結果を利用した。これらを元にしてズーム操作推定アルゴリズムを以下の手順で作成する。

- (1) センサデータを図 1 のように固定長の時間幅で分割する。
- (2) 分割した区間ごとに特徴量を計算する。
- (3) 区間内で半分以上着目されていれば T、そうでなければ F とラベルを付ける。
- (4) これにより、図 1 のような学習データが作成される。赤枠はユーザが着目した箇所、データセットの青丸は特徴量、一番右がラベル示している。
- (5) T / F を目的変数に、それ以外の特徴量を説明変数として教師あり機械学習を行う。その結果、センサデータから着目点を出力するアルゴリズムができる。

この作成した学習アルゴリズムを利用することにより、新たなセンサデータに対して自動的に着目すべき点を提示することができる。

本稿では、センサデータとして本研究室が行なっている介護施設での行動分析実験の際に得られている 2 種のセンサデータを利用して実験と評価を行なった。可視化された 2 種のセンサデータに対して理系大学院生を対象とした被験者が着目した箇所に赤枠を付けてもらう。その時に赤枠の座標がどの時間帯にあるかを取得する。そして手順に従ってデータセットを作成し、教師あり機械学習を行う。手順 (5) の教師あり機械学習は今回、ランダムフォレスト法を利用した。

4. 評価

本稿では以下の 4 つについて評価を行う。

- (1) 着目点を推定できたか
- (2) どの特徴量の重要度が高いか
- (3) 推定結果を実際に可視化した場合に見やすいか

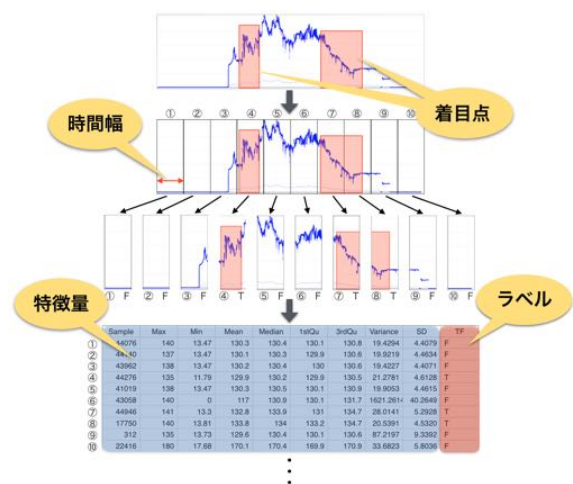


図 1 データセットの作成方法の例

- (4) ある可視化手法の学習アルゴリズムを利用して他の可視化手法でも推定可能か

(1) に関しては、実際にユーザが着目した箇所を推定できているか評価するために行う。(2) に関しては、ユーザが可視化されたデータにどのような特徴がある場合に着目しているか評価する。そのためにランダムフォレストから得られる特徴量の重要度から分析する。(3) に関しては、着目点の推定結果から可視化した場合にその可視化が有用であるかを評価する。そのために実際に着目点を推定して可視化を行う。(4) に関しては、各センサごとにある 1 種の可視化手法の学習アルゴリズムを利用して他方の可視化手法にも適用可能か評価する。具体的に照度センサを箱ひげ図で可視化した場合に作成した学習アルゴリズムを棒グラフの場合でも推定可能か、また逆の場合も推定可能であるかを評価する。

4.1 介護施設でのセンシングデータ

介護施設でのセンシング内容は、スマートフォンの照度センサ及びセンサタグの加速度センサを利用して 100 ms 間隔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。本稿ではここで得られた照度センサと加速度センサそれぞれ 50 日分のセンサデータを利用した。

4.2 評価システム

実験を行うにあたり、図 2 のような Web 上で動作する評価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。画面の構成は画面中央

部にセンサデータを可視化したものを表示する。この可視化されたものに対して赤枠で着目点に印をつける。着目点が複数ある場合、画面下部にある「枠の追加」のボタンを押すことによって赤枠を追加することができる。全ての着目点に赤枠を付け終わると「次へ進む」のボタンを押すことで次のデータへ進むことができる。具体的な実験方法については次項で述べる。

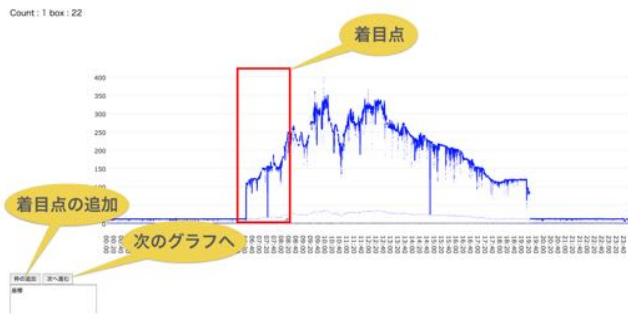


図 2 使用する評価システム

4.3 評価実験

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加速度センサの2種類とした。照度センサは1日分のデータ、加速度センサは1時間分のデータをランダムに可視化した。その際、センサごとに考えられる2種の可視化手法を用いて可視化を行った。照度センサは図3と図4のように1分おきの箱ひげ図と10分おきの平均値を表した棒グラフの2種、加速度センサは図5と図6のようにX軸、Y軸、Z軸別々の場合と3軸合成値の場合の2種を折れ線グラフで可視化した。この可視化されたグラフに対して、被験者に赤枠で着目点に印を付けてもらう。このとき、3節(1)での固定長の時間幅を照度センサは1時間ごと、加速度センサは2分ごとに分割した。さらに、3節(2)での特徴量の計算は、「最大値(Max)」、「最小値(Min)」、「平均(Mean)」、「中央値(Median)」、「第1四分位数(1stQu)」、「第3四分位数(3rdQu)」、「分散(Var)」、「標準偏差(Sd)」、「データ数(Sample)」の9種類とした。3節(3)では赤枠の座標を取得してTまたはFを判断する。この作業を1サイクルとして、23歳男性の理系大学院生5人の被験者に1種の可視化手法につき50サイクル行ってもらい計4種類分のデータを収集した。これにより3節(4)で示すデータセットが作成することができる。この作成したデータセットをランダムに半分に分け、それぞれを学習データとテストデータとして、3節(5)で示すようにT/Fを目的変数に、それ以外の特徴量を説明変数としてランダムフォレスト法で学習を行う。本稿では具体的にR言語のrandomForestパッケージを利用した。

この3節(5)によって作成したアルゴリズムを利用して

評価項目(1)に関して評価を行う。各センサの各可視化手法においてテストデータを用いて実際に着目点を推定することができるか、Precision, Recall, F値を用いて評価を行う。さらにランダムフォレスト法を適用した際に得られた特徴量の重要度から評価項目(2)の評価を行う。評価項目(3)に関しては項目(1)で着目点を推定したものを実際に可視化し評価を行う。評価項目(4)に関しては具体的に、照度センサを箱ひげ図で可視化した場合に作成したアルゴリズムを棒グラフの場合でも推定可能か、また逆の場合も推定可能であるかを評価する。また、加速度センサの場合も同様にして評価を行う。



図 3 照度センサの可視化の例(箱ひげ図)

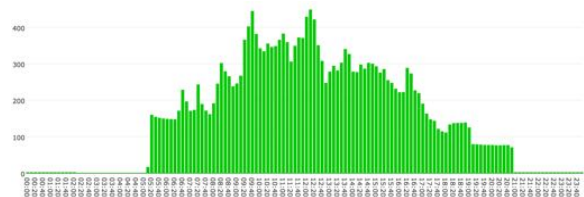


図 4 照度センサの可視化の例(棒グラフ)

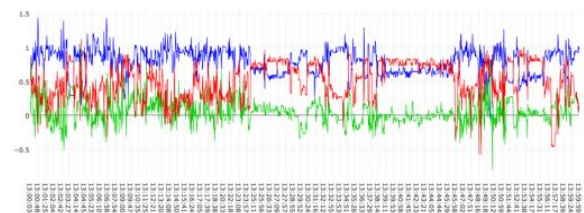


図 5 加速度センサの可視化の例(X軸、Y軸、Z軸別)

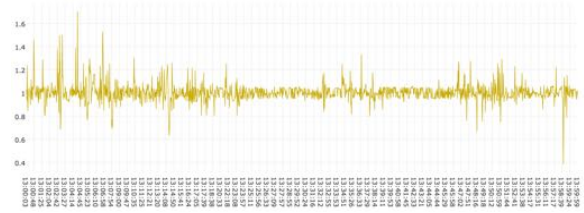


図 6 加速度センサの可視化の例(3軸の合成値)

4.4 評価結果

まず評価項目 (1) の各可視化手法においてテストデータを用いて実際に着目点を推定することができるか評価を行う。表 1 は各センサの各可視化手法ごとの Precision, Recall, F 値の平均値を示す。照度センサを箱ひげ図で可視化した場合の F 値は 81.1%, 棒グラフで可視化した場合の F 値は 83.2%と高い精度を得ることができた。しかし、加速度センサの X 軸, Y 軸, Z 軸の 3 軸別々で可視化した場合の F 値は 61.8%, 3 軸合成値で可視化した場合の F 値は 69.0%とチャンスレベルの 50%よりも高い精度を得ることはできたが、照度センサの 2 種の可視化手法と比較すると低い精度となった。

表 1 各可視化手法ごとの精度

センサ	可視化手法	Precision	Recall	F 値
照度	箱ひげ図	81.1%	81.1%	81.1%
	棒グラフ	83.3%	83.2%	83.2%
加速度	折れ線 (X,Y,Z)	61.8%	62.0%	61.8%
	折れ線 (3 軸合成値)	68.5%	71.8%	69.0%

次に評価項目 (2) についてデータにどのような特徴がある場合に着目したのか特徴量の重要度から評価を行う。図 7-10 は各可視化手法でランダムフォレスト法を適用した際に得られた特徴量の重要度を示す。照度センサの 2 種の可視化手法の場合、「標準偏差」「分散」が特に重要度が高い特徴量であった。次いで「最大値」と「平均値」が重要度が高いという結果になった。加速度センサの 3 軸別々に可視化した場合、X 軸の「最小値」「最大値」「分散」「標準偏差」が特に重要度が高い特徴量であった。また、3 軸合成値を可視化した場合も X 軸の「最大値」「標準偏差」「分散」「最小値」が重要度が高くなっており、加速度センサの 2 種とも X 軸の特徴量が特に重要度が高いという結果になった。

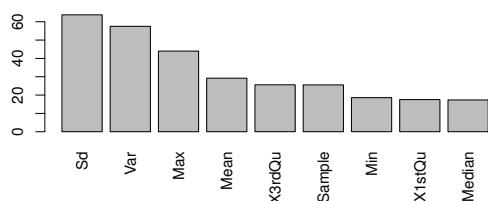


図 7 箱ひげ図の場合の特徴量の重要度

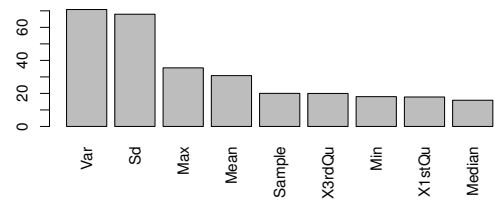


図 8 棒グラフの場合の特徴量の重要度

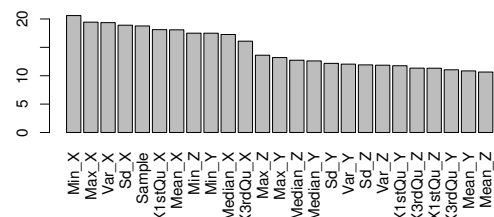


図 9 3 軸別々で可視化した場合の特徴量の重要度

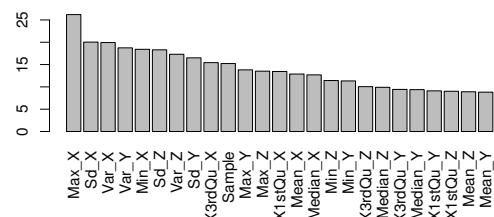


図 10 3 軸合成値で可視化した場合の特徴量の重要度

次に評価項目 (3) の実際に推定結果を可視化した場合について評価を行う。図 11-14 は評価方法 (1) によって着目点を推定した結果であり、上部の赤枠は推定箇所、下部は赤枠をズームして可視化したものである。照度センサを箱ひげ図と棒グラフで可視化した場合、図に示すように推定された着目範囲は両者ともほぼ同じ場所であった。また、推定箇所のみを可視化すると変化が明瞭な箇所がズームされて可視化されているため観察しやすくなった。しかし、加速度センサの場合、可視化する手法によって推定箇所に大きな違いが見られた。3 軸別々に可視化した場合、推定箇所は図 13 に示すように小さい範囲であった。しかし、3 軸合成値の場合、推定箇所は図 14 に示すようにほとんどの範囲が着目点と推定された。加速度の場合、被験者によって着目点が変わり、さらに同じ被験者でも可視化手法ごとに着目点が変わっていたためこのような違いが見られたと考えられる。また、推定箇所をズームして可視化したところ、照度センサの場合と同じく観察しやすくなった。

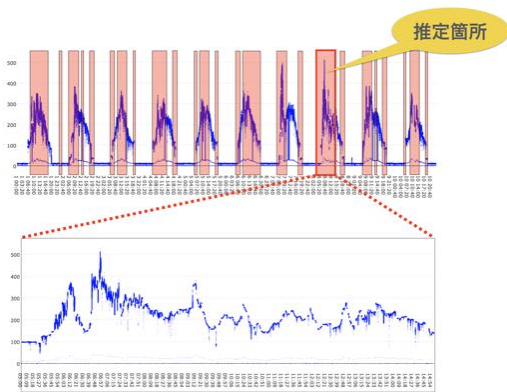


図 11 箱ひげ図の場合の推定結果を可視化した例

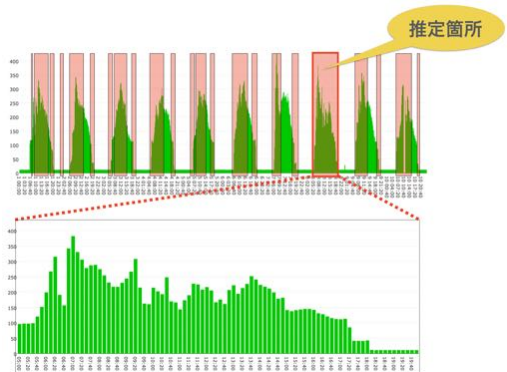


図 12 棒グラフの場合の推定結果を可視化した例

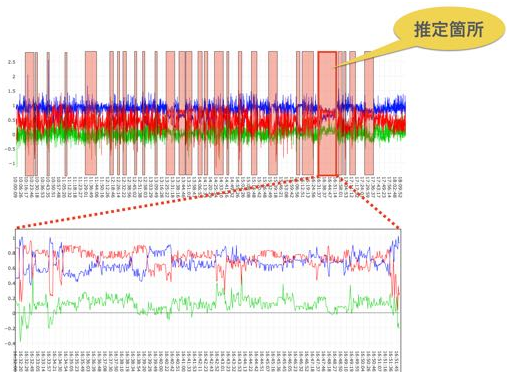


図 13 3 軸別々で可視化した場合の推定結果を可視化した例

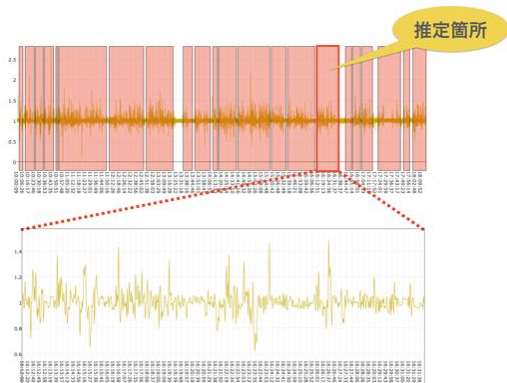


図 14 3 軸合成値で可視化した場合の推定結果を可視化した例

次に評価項目 (4) のある可視化手法で学習したものを他の可視化手法に適用できるか評価を行う。表 2 は別の可視化手法のアルゴリズムを適用したときの Precision, Recall, F 値の平均値を示す。1 行目は照度センサの棒グラフで学習したアルゴリズムを箱ひげ図に適用した結果を示しており、F 値は 83.3% と高い精度を得ることができた。2 行目は 1 行目の逆のケースを示しており、こちらも F 値は 83.2% と高い精度となった。3 行目は加速度センサの 3 軸合成値の場合で学習したアルゴリズムを 3 軸別々に可視化したものに適用した結果を示している、F 値は 55.4% とチャンスレベルの 50% よりも高い精度だが評価方法 (1) における精度よりも低くなっている。4 行目は 3 行目の逆のケースを示しており、こちらも F 値は 55.4% とチャンスレベルよりも高くはなっているが評価方法 (1) の場合よりも低くなっていることがわかる。

表 2 別の可視化手法のアルゴリズムを適用したときの精度

センサ	学習	テスト	Precision	Recall	F 値
照度	棒グラフ	箱ひげ図	83.3%	83.3%	83.3%
	箱ひげ図	棒グラフ	83.3%	83.2%	83.2%
加速度	3 軸合成値	X,Y,Z	55.3%	55.6%	55.4%
	X,Y,Z	3 軸合成値	55.6%	55.3%	55.4%

5. 考察

本稿では、照度センサと加速度センサそれぞれで 2 種の違う可視化を行い、ユーザの着目点を目的変数に、それ以外の特徴量を説明変数としてランダムフォレスト法を利用して評価を行なった。その結果、照度センサの場合は箱ひげ図と棒グラフのどちらも高い精度を得ることができ、「分散」と「標準偏差」が重要度の高い特徴量であった。これは被験者が可視化されているデータの変化が明瞭な部分に着目しているためと考えられる。実際に被験者がデータのどの時間帯に着目したのかを示したものが図 15 である。被験者個々で見た場合、被験者 C が箱ひげ図では 1 箇所に着目したのに対し、棒グラフでは 2 箇所に着目している。それ以外の被験者は両者ともほぼ同じ場所に着目しており、被験者全員がデータに変化がない時間帯には着目せずに変化が明瞭な部分に着目していることがわかる。そのため、箱ひげ図で学習したアルゴリズムを棒グラフに適用した場合、またその逆の場合も精度も落ちることなく高い精度となった。さらに、推定結果を可視化したところ、箱ひげ図と棒グラフのどちらの場合も変化が大きい時間帯のみがズームされたため、どのように変化しているのかわかりやすくなった。

次に加速度センサは照度センサと比較すると低い精度となった。これは図 16 に示すように、被験者によって着目している時間帯が異なっているため、照度センサよりも精度が落ちたと考えられる。3 軸別々の場合は、被験者 A,E

は全体的に着目しているのに対し、被験者 B,C は 1 箇所のごく一部しか着目していない。さらに、3 軸別々で可視化した場合と 3 軸合成値で可視化した場合を比較しても着目点が異なっていることがわかる。一方の可視化手法で学習したアルゴリズムを他方の可視化手法に適用した場合の精度が低かった事実と一貫する。また、推定結果を実際に可視化したところ、3 軸別々の場合は着目点と推定された時間帯はわずかであったが、3 軸合成値の場合はほとんどを着目点と推定した。

このように、照度センサのようなデータの変化が明瞭な場合には着目点を高精度で推定することが可能となる。さらに可視化手法が類似している場合には、1 つの学習されたアルゴリズムで複数の可視化手法に適用できることが確認できた。

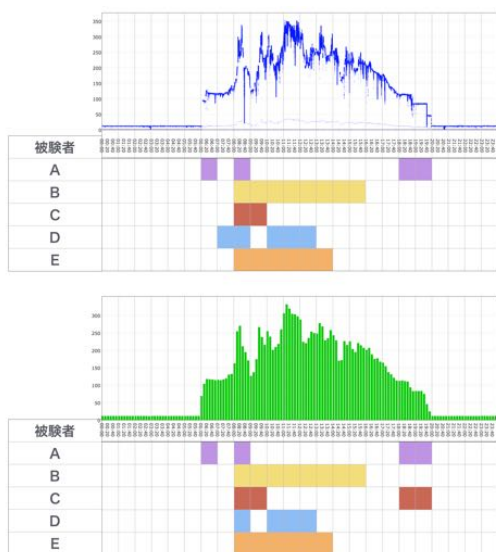


図 15 被験者ごとの着目点 (照度センサ)

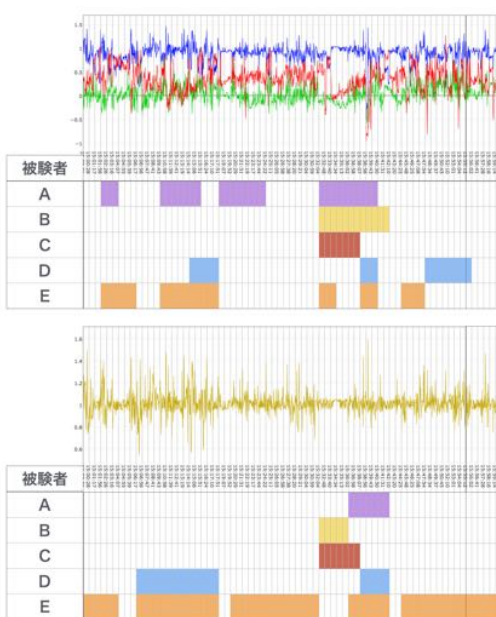


図 16 被験者ごとの着目点 (加速度センサ)

6. まとめ

本研究では、センサデータを自動的に適切な可視化を実現する手法を提案し、評価を行なった。その結果、照度センサなどのデータの変化が明瞭な場合には、可視化手法が異なってもほとんどのユーザはデータの変化が大きい部分に着目することがわかった。そのため高い精度で着目点推定が可能であった。しかし、加速度センサなどのデータが常に変化し続けている場合には、ユーザによって着目点が異なり、同じユーザでも可視化する手法によって着目点が異なることがわかった。したがって、照度センサなどのデータの変化が明瞭な場合に比べると精度が落ちてしまう。今後は、加速度データに対して、推定精度を向上可能な可視化手法や特徴量を探索する必要がある。また将来的には、他のセンサデータや可視化手法で実験を行い、1 つの学習アルゴリズムでさまざまなセンサデータや可視化手法に適用してユーザ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.
- [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, pp. 416-425, 2016.
- [3] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, J. Heer, "Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations", IEEE Trans. Visualization & Comp. Graphics, 2015.
- [4] W. Javed and N. Elmqvist, "Stack zooming for multi-focus interaction in time-series data visualization," in Proceedings of the IEEE Pacific Visualization Symposium, 2010, pp. 33-40.
- [5] 井元麻衣子, 伊藤貴之, "SAX 法による局所パターン抽出を導入した時系列データの三次元可視化の一手法", 芸術科学会論文誌, Vol. 10, No. 3, pp. 179-191, 2011.