



環境・生体ログからの行動パターン解析における 分類精度の向上

1615033 沼田 賢一

情報基盤工学講座 指導教員 奥原浩之

要約

本研究では、ログの行動パターンの類似性・イベント検出の向上を考える。先行研究において、画像認識APIを用いてログで取得したカメラ画像をテキスト化し、その他のセンサデータを含めてログの解析を行うものがある。しかし、テキスト化した文字をそのまま解析しても分類精度が高いとは言えない。そこで、テキストをベクトル化させ、さらにログに教師データを付けることで従来の分類精度の向上を図る。

キーワード： ライフログ、画像認識、Word2vec、機械学習

1 はじめに

現代、多くの人がスマートフォンやウェアラブルデバイスを持ち歩くことが一般的であり、急速な情報技術の発達から、個人の生活や行動をデータとして取得、記録することが可能となっている。

このようなスマートフォンやウェアラブルデバイスを使用して取得して得られる人間の活動の記録のデータをログデータといふ。ログデータは解析を行うことで、個人の健康管理に活かしたり、ビジネスとしてターゲティング広告にも使われ社会に活かしたりできると考えられている。

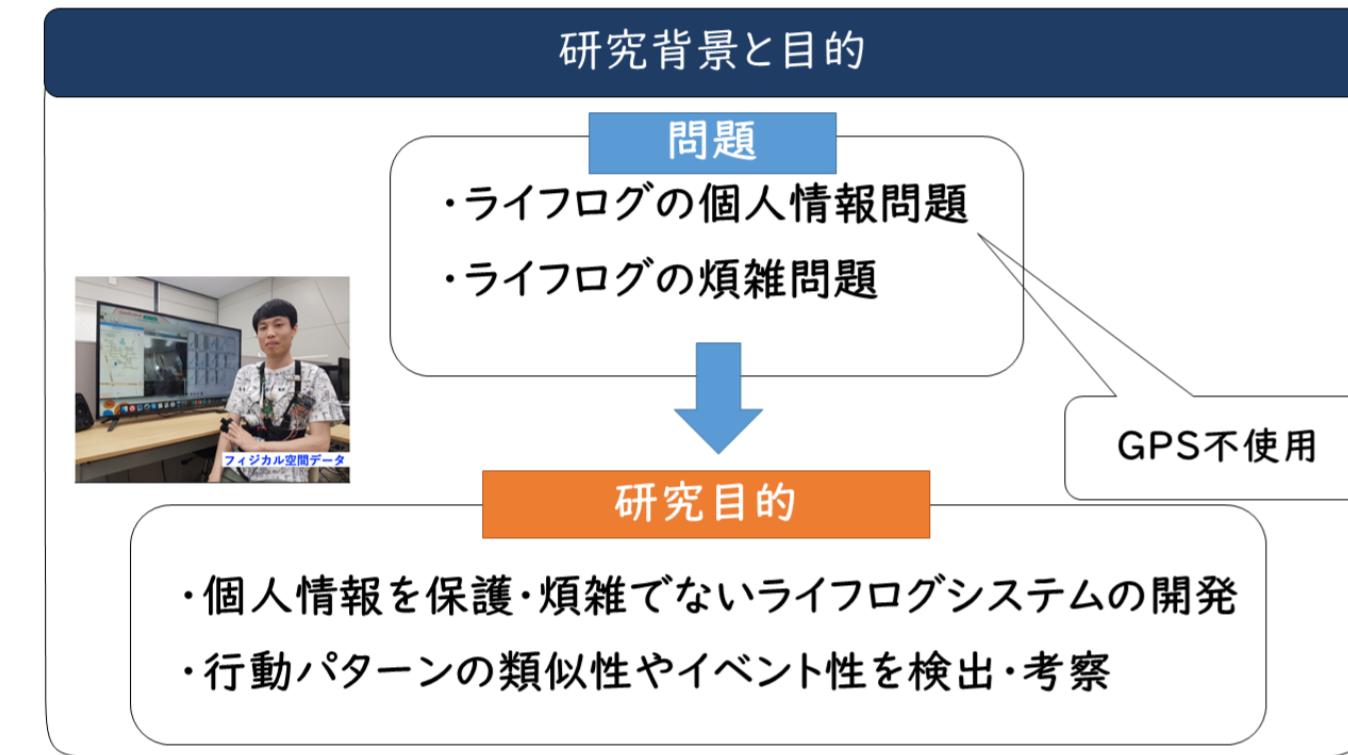


図1 研究背景と目的

2 環境・生体ログ

2.1 従来研究の紹介 先行研究では、KH CoderとR言語を用いて、多変量解析からログデータの行動を識別し、自己組織化マップから時系列を可視化する。入力層と出力層の2つに分かれ、入力層から出力層への入力を x 、出力層のニューロンの番号を i 、参照ベクトルを m_i と定義し、入力ベクトルと各ニューロンの参照ベクトルとの距離で出力層のニューロンを競合させ、勝者ニューロン c の参照ベクトルと入力ベクトルを近づける学習を行っている。

従来研究の例

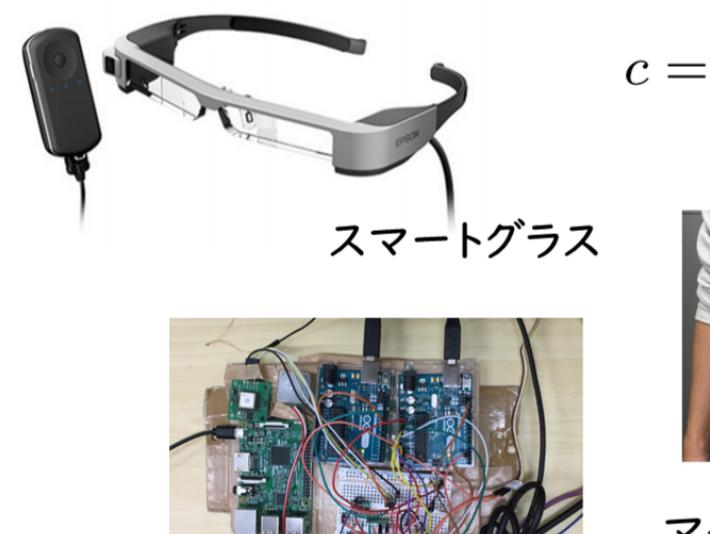


図2 従来研究の例

2.2 従来の分析手法

先行研究の手法については、データ取得部と行動識別部から構成されるシステムとなっている。データ取得部で、視界情報をテキストに変換しながらログデータを取得する。

データを取得するために、カメラ画像を一定時間ごとに自動で取得し、画像認識APIを用いて画像情報をテキストに変換・記録するアプリケーションを開発し、行動識別部で、記録したデータを整理し、多変量解析を用いて解析・比較を行うことで、行動を識別し周期性を検出していた。

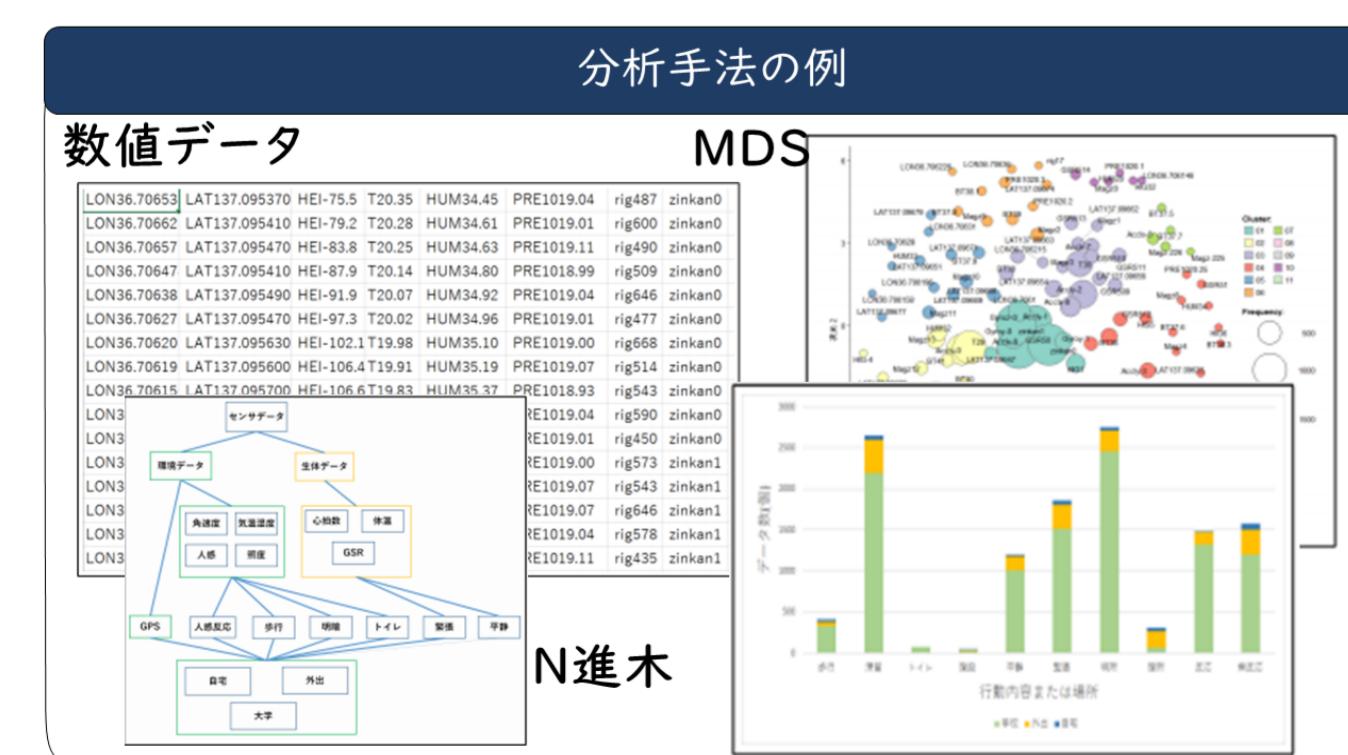


図3 分析手法の例

2.3 事象抽出によるパターン検出

ログデータに対して多変量解析を行い、解析結果の読み取り・比較を行うことで一定時間内の行動識別を行い、ログデータの類似性やイベント性を検出できると考える。本研究では、ログデータの類似性とは、多変量解析によるクラスターの分かれ方から導き出せる行動や、プロットの関係性、SOMであらわされる時系列が類似している場合類似性があると考える。一方で、ログデータから特徴的なイベント性を検出した場合、イベント日であることを検出できたり、平常日とは違うという危険を察知したりできる。

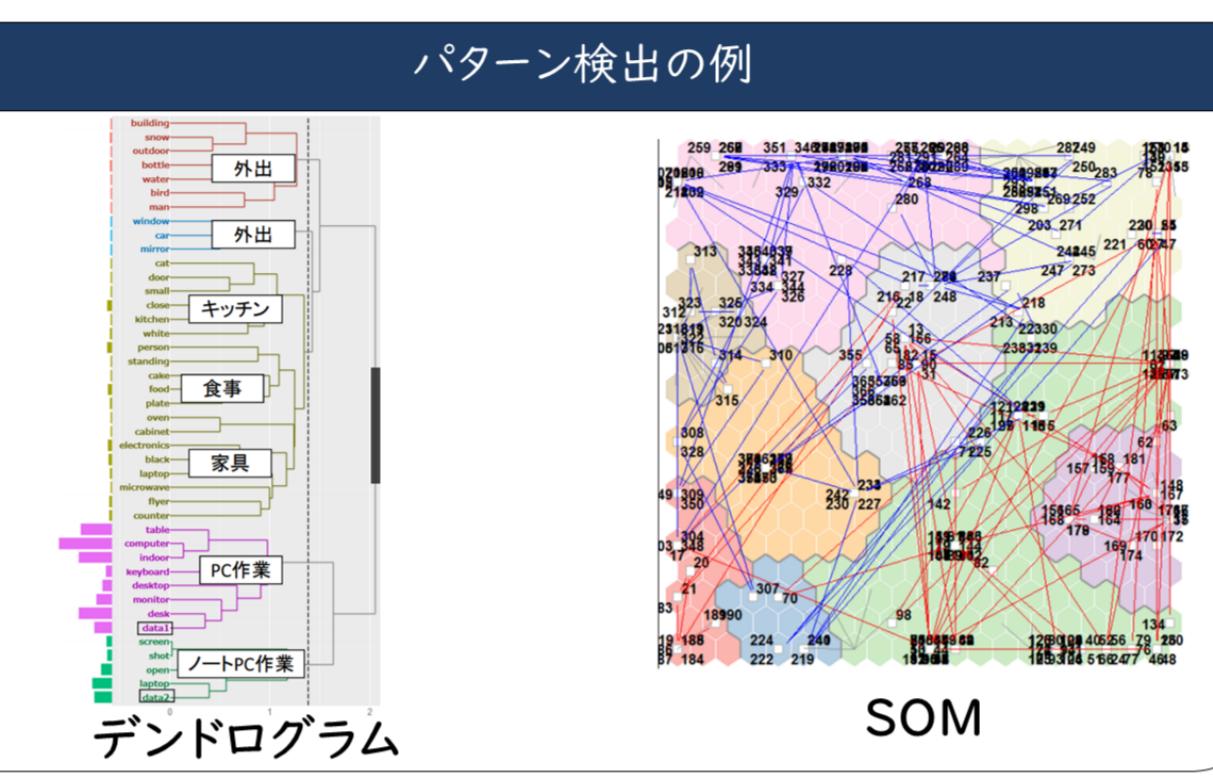


図4 パターン検出の例

3 行動パターン解析

3.1 入力信号によるラベル付け

ログデータの収集時に画像のテキスト変換と同様に、同時に音声マイクを用いて、現在の行動を認識できるような簡易的なテキスト変換APIを用いて定期的に入力信号として収集することでログデータの解析する際に入力信号を教師データとして含めることで分類精度を向上させる。

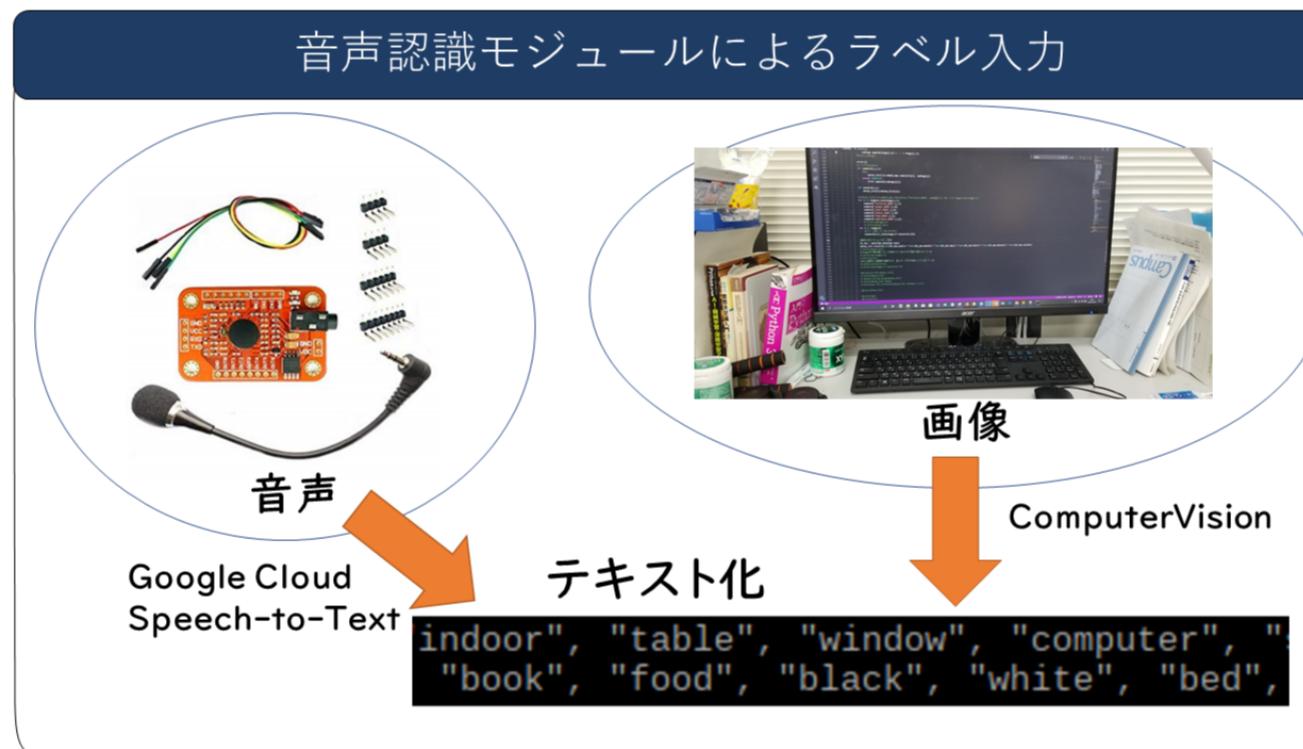


図5 音声認識モジュールによるラベル入力

3.2 テキストデータのベクトル化

Word2vecはMikolovらにより発表された単語群のベクトル化手法である。Word2vecを自然言語処理の分野で応用した例に、日本語動詞・形容詞に関する類似度データセットの構築やイベント情報の分類など数多くの利用例がある。本研究では、カメラから取得した画像を画像認識APIを通してテキスト化させたものに対してWord2vecを用いてベクトル化させる。そして家具、動作、食べ物といった6つの基準を設けそれらとの類似度の値を求める。

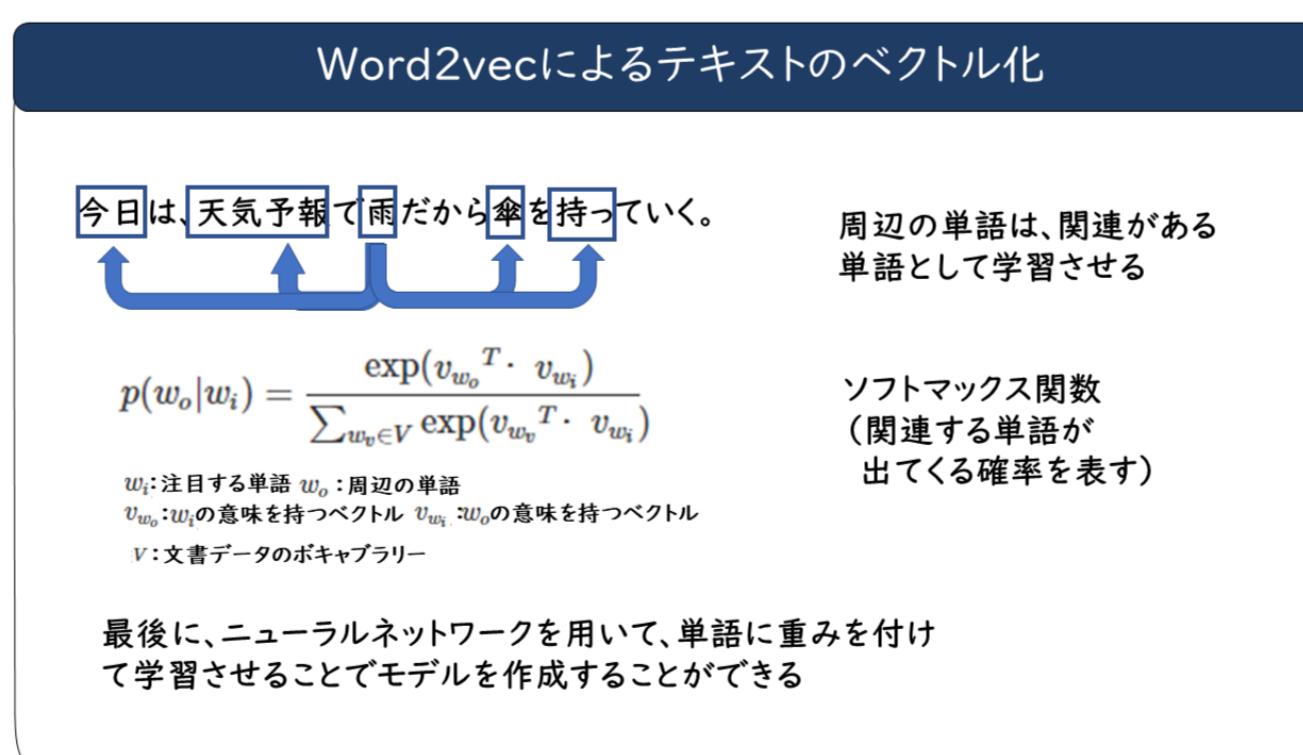


図6 Word2vecによるテキストのベクトル化

り方と比較することで分類精度が向上していることを確かめる。

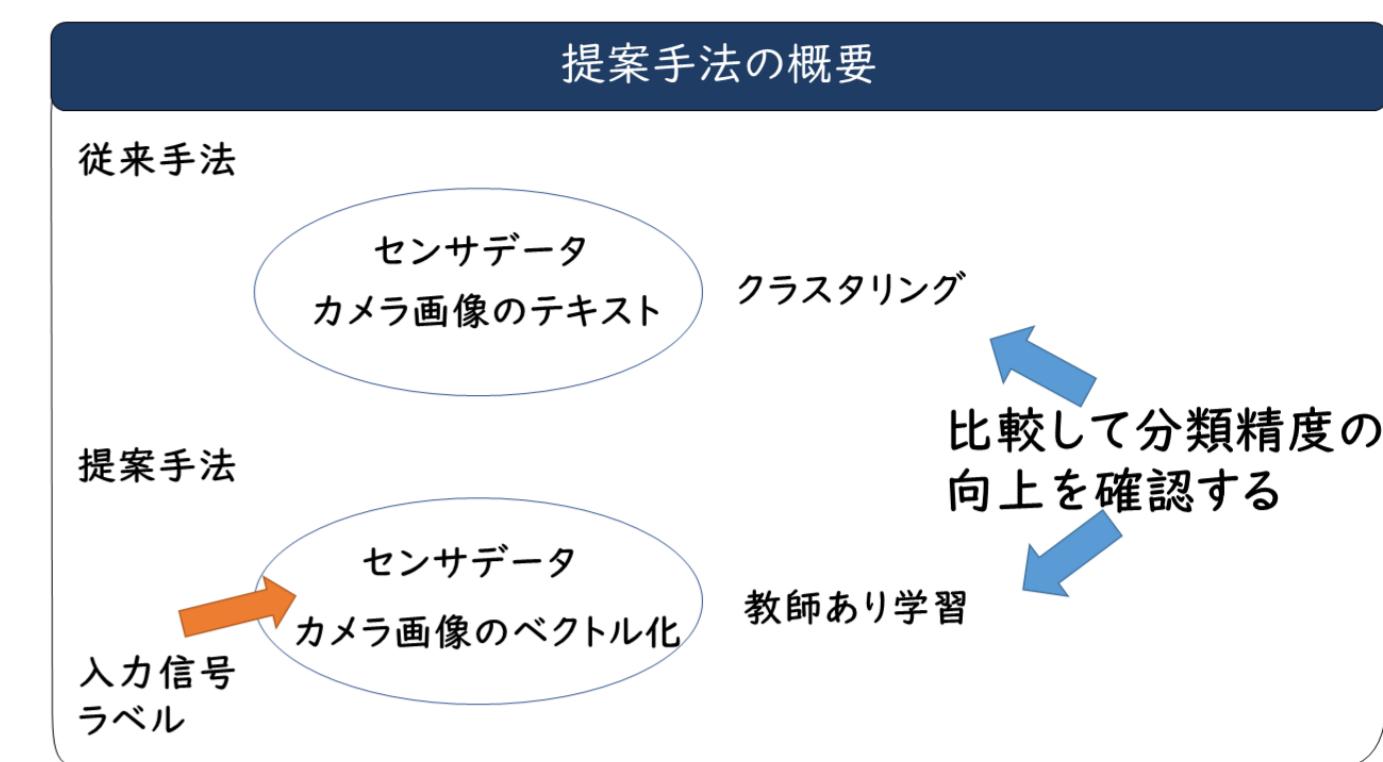


図7 提案手法の概要

4 数値実験ならびに考察

データを構成する行動の検出、類似する行動とそうでないイベント性のある行動を検出することができた。また、自己組織化マップの解析から、同じ行動でも視界に写る物体の違いから行動の類似性やイベント性を検出できた。この結果から、同じ行動でも使用する場所や物体の変化によって別行動として認識させることができたためGPSを使用せず、ログデータに位置情報を付加できると考えられる。

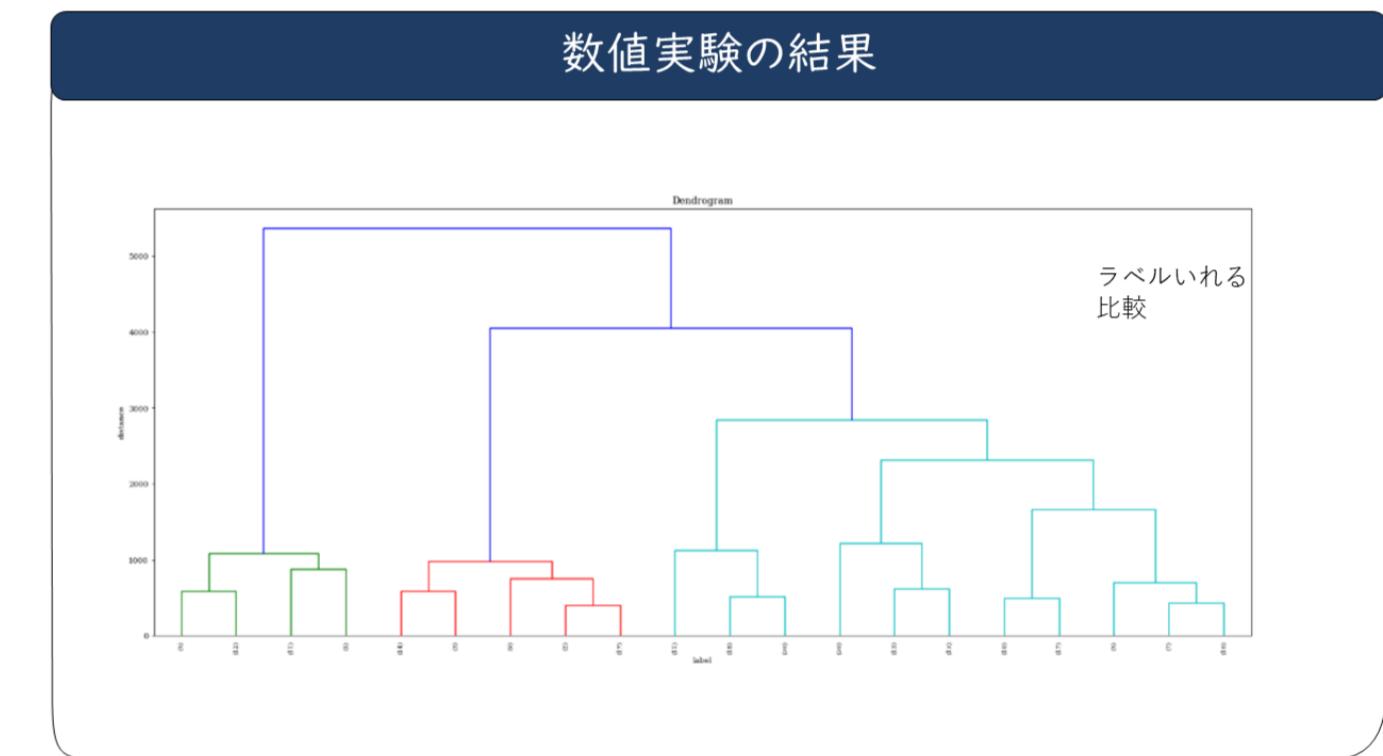


図8 数値実験の結果

5 おわりに

個人情報保護に着目したログデータから類似性やイベント性を検出できた。今後の課題として、ユーザーの取得タイミングを組み込んだログデータ取得アプリケーションの開発、長時間のログデータ取得にデバイスが耐えうるようプログラムを改善することが挙げられる。

参考文献

- [1] “環境認識ログからの行動パターン解析による類似性・イベント検出”
- [2] “第2章 ネットワーク構造つながりを見る” <http://www.econ.tohoku.ac.jp/ksuzuki/teaching/ch2.pdf>
- [3] “建設工事における総括工程計画モデルの開発研究”
- [4] 飯田耕司, “不確実性への挑戦・意思決定分析の理論”
- [5] 現場コミュニケーションアプリ「Kizuku(キズク)」
- [6] Hideki Katagiri, Interactive multiobjective fuzzy random linear programming: Maximization of possibility and probability
- [7] M. Sakawa, I. Nishizaki, H. Katagiri, Fuzzy Stochastic Multiobjective Programming, Springer, 2011
- [8] 椎名孝之, “確率計画法”, 朝倉書店, 2015