

要約

本研究では、ライフ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を用いてライフ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ライフ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、さらにライフ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ライフログデータという。ライフ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。

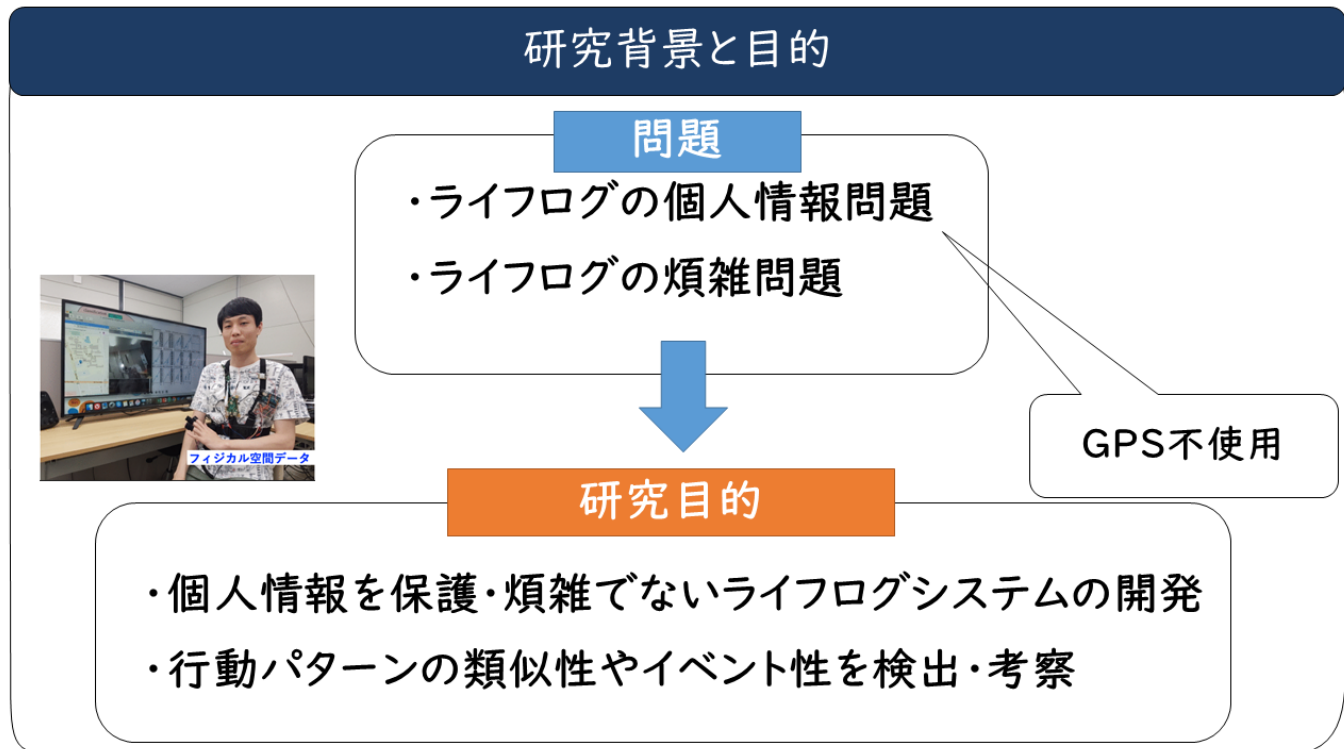


図1 研究背景と目的

2 環境・生体ライフログ

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, 多変量解析からライフログデータの行動を識別し, ライフログデータの類似性やイベント性を検出したものがある [1]. また, 人物写真に対してクラスタリングを行いその精度と有効性を検証したものがある [4]. さらに, 断片的なライフログデータの解析の支援を行うためのフレームワークを作るような研究がある [5].



図2 従来研究の例

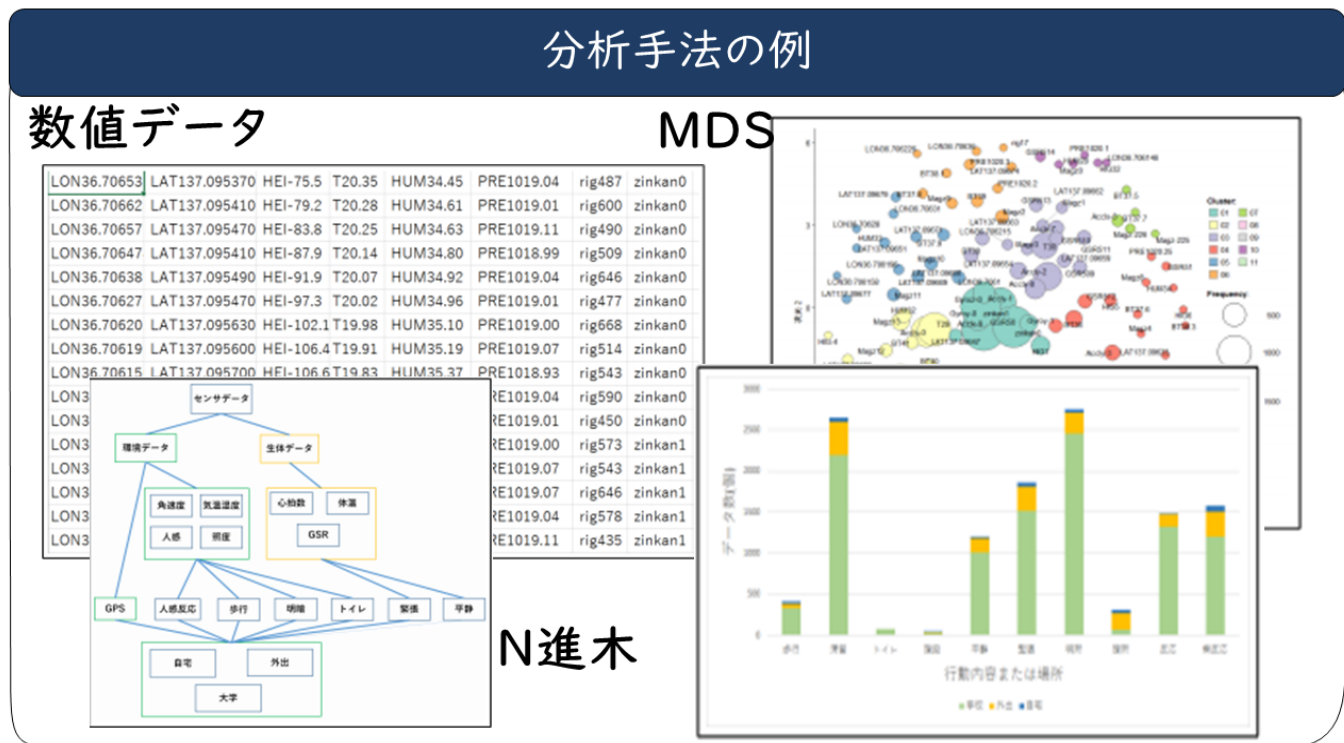


図3 分析手法の例

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ライフ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 [2].

2.3 事象抽出によるパターン検出

ライフ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, ライフ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, ライフ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える。一方で, ライフ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, 平日日とは違うという危険を察知したりできる。

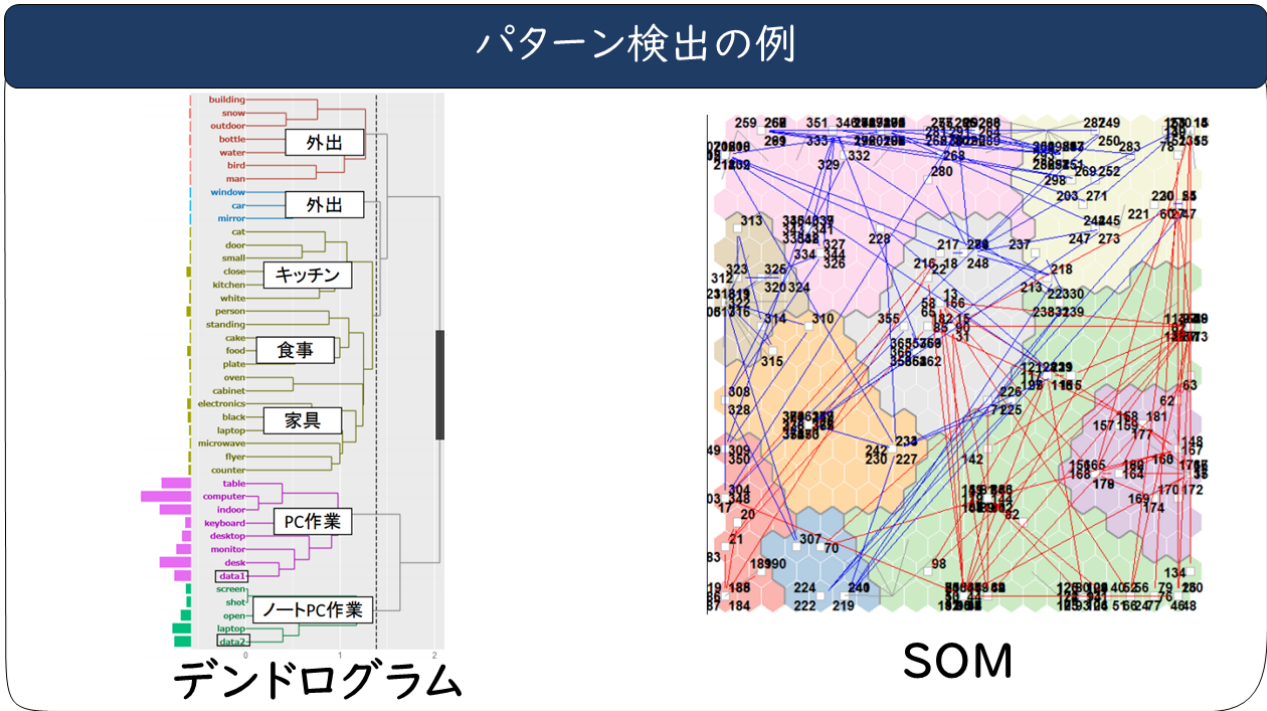


図4 パターン検出の例

3 行動パターン解析

3.1 入力信号によるラベル付け

ライフ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に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ライフ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。

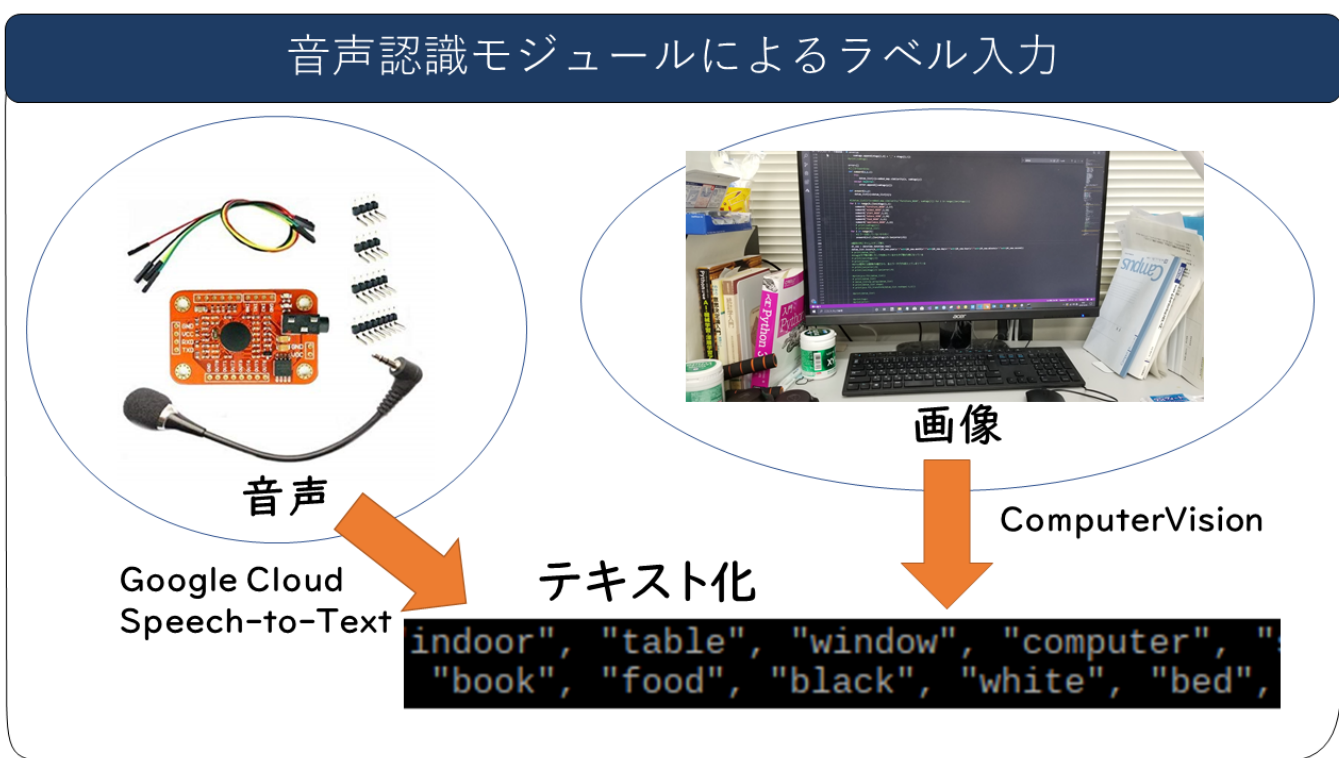


図5 音声認識モジュールによるラベル入力

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。

