

- 1. はじめに
- 2. ライフログと各種センサ
- 3. 提案手法
- 4. 数値実験ならびに考察
- 5. 今後の課題

# 環境・生体ライフログからの 画像・音声分析と 単語ベクトルによる行動識別

Behavior Identification

Using Image and Speech Analysis and Word Vectors  
from Environmental and Biological Life Logs

Keniti Numata

Graduate School of Information Engineering, Toyama Prefectural University  
t615033@st.pu-toyama.ac.jp

F108, 9:45-10:00 Thursday., February 17, 2020,  
Toyama Prefectural University.

# 1.1 本研究の背景

2/16

## 背景

近年、ウェアラブルデバイスやスマートフォンが普及し、一般的にライフログを取得できるようになった。また、デバイスを用いて集めたライフログデータをもとに健康管理を行うアプリケーションが存在する。

## ライフログデータについて

- 1 ライフログとは、人間の生活（life）をデジタルデータとして記録（log）する技術である
- 2 ライフログデータとは、個人の行動履歴に基づいて生み出されるビックデータである
- 3 健康管理や、安否確認に利用できる

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

## 1.2 本研究の目的

3/16

### ライフログの問題点

#### 1 煩雑問題

意識的にライフログデータを取得する必要があり手間がかかる

#### 2 多様化問題

さまざまなアプリケーションが存在するが、それぞれで扱うデータが異なるため、ユーザーが正確に管理するのが難しい

### 本研究の目的

- ・ ライフログデータの収集機器を開発
- ・ 個人のライフログデータを収集
- ・ ライフログの問題点を解決し行動識別ができるシステムの作成

## 2.1 開発したライフログ収集機器

4/16

### 環境・生体データ収集機器

マイコン (Arduino と Raspberry Pi 3.0) とセンサ類を組み合わせ  
て独自のライフログの測定機器を開発する。

#### 収集データ一覧

- 環境センサ「GPS(緯度, 経度), 温度, 湿度, 気圧, 照度, 人感, 加速度 (3 軸), 角速度 (3 軸), 磁気コンパス (3 軸), カメラ, マイク」
- 生体センサ「体温, 心拍, ガルバニック皮膚反応」

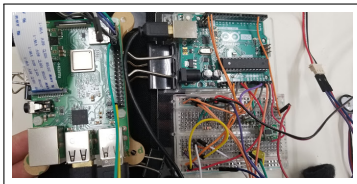


Figure 1: データ収集機器

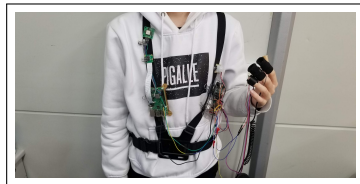


Figure 2: 装着時

## 2.2 無線によるセンサデータ収集

5/16

### クラウドの使用

サーバー PC で行動識別を行うため収集したセンサデータを、ソケット通信でサーバー PC に送信する。

Wi-Fi を切り替えたときに、IP アドレスが変わる  
→クラウドを経由してサーバー PC に送信した。

### Google スプレッドシート

使ったクラウドは、Google スプレッドシートである。

Google スプレッドシートを外部から操作するために Google App Script を使う。

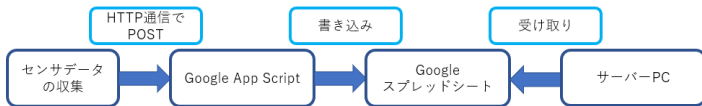


Figure 3: ライフログデータの流れ

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

## 3.1 提案手法

6/16

- 1. はじめに
- 2. ライフログと各種センサ
- 3. 提案手法
- 4. 数値実験ならびに考察
- 5. 今後の課題

### 従来の行動識別の問題点

ライフログデータ行動識別の関連研究では、KH Coder を用いて分析する研究 [1] がある。

また、カメラの静止画像を API を使って写っているモノをテキストとして取り出して他のセンサデータの数値と合わせて KH Coder を用いて分析する研究がある。

KH Coder は、テキスト型の分析に向けたソフトであり、分析するためにパラメータを毎回設定する必要がある。

### 行動識別のための提案手法

- テキストデータのベクトル化
- 入力信号によるラベル付け
- クラスタ分析による行動識別

[1] 樋口耕一, “テキスト型データの計量的分析: 2つのアプローチの峻別と統合”, 理論と方法, Vol.19, No. 1, pp. 101–115, 2004.

## 3.2 テキストデータのベクトル化

7/16

テキストデータのベクトル化をするために、画像認識 API と word2vec を用いた。

### 画像認識 API

画像認識 API とは、大量の画像で機械学習したモデルを用いて画像の特徴を出力する API である。使用する API は、Computer Vision API である。

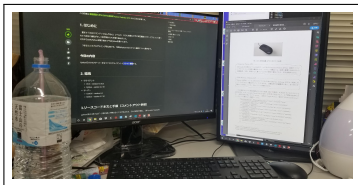


Figure 4: サンプル画像

```
{ "tags": [ "computer", "indoor", "desk", "table", "monitor",
"keyboard", "electronics", "sitting", "bottle", "laptop", "black",
"desktop", "wooden", "mouse", "white", "phone" ], "captions": [ {
"text": "a desktop computer monitor sitting on top of a desk",
"confidence": 0.8872777 } ] }
```

Figure 5: 出力結果

## 3.2 テキストデータのベクトル化

8/16

### word2vec

word2vec とは、分散表現のモデルを使って単語をベクトル化する手法である。モデルは、事前に学習済みのモデルを使用する。

### 類似度

word2vec を用いると単語同士の類似度を計算できる。単語をそのままベクトル化すると多次元になるので、基準の単語をきめて、その単語との類似度で単語から数値にする。基準の単語は furniture, animal, plant, behave, food, appliance の 6 つを設定した。

	基準	furniture	animal	plant	behave	food	appliance
食事	テキスト	person,food,plate,table,holding,eating,woman,sitting,young,white,filled,hot,little,donut,sandwich,meat,rice,dog					
	数値化後	0.21928	0.242008	0.154744	0.149327	0.304002	0.15837
パソコン作業	テキスト	indoor,computer,desk,table>window,sitting,laptop,monitor,keyboard,office,white,room,desktop,refrigerator,bed					
	数値化後	0.287702	0.162321	0.133619	0.15251	0.165043	0.275778

Figure 6: 行動の違いによる数値の比較

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題



## 3.2 テキストデータのベクトル化

9/16

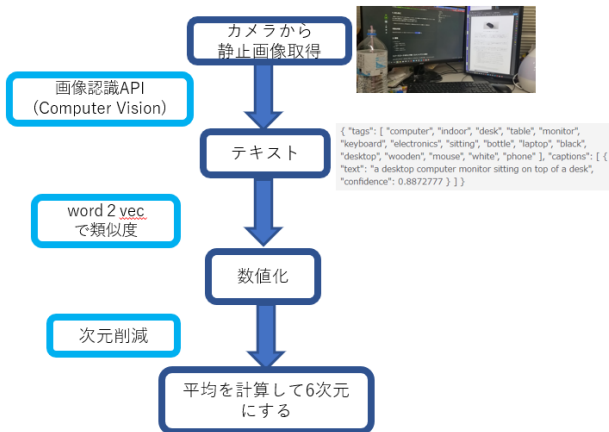


Figure 7: テキストデータのベクトル化

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

## 3.3 入力信号によるラベル付け

10/16

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

### 音声認識 API

音声認識 API とは、録音された音声をテキストに変換してくれる API である。

音声認識 API を使って取得したテキストをクラスター分析のデンドログラムのラベルに使用した。

また現在の行動の識別をするために、行動がどのクラスタにあるかデンドログラムで見るために、最後に取得したセンサデータのラベルには「最新」というラベルつけた。

## 3.4 クラスター分析による行動識別

11/16

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

### クラスター分析

クラスター分析とは、大きな集団の中から似た者同士を集めてグループに分ける統計的な分析手法である。クラスター分析には、階層的手法と非階層的手法の2つの手法がある。非階層的手法は、クラスタ数を決める必要がある。

ライフログを取得するとき行動の種類は決まっていないので今回は、階層的手法を用いた。

また、クラスター分析の結果はデンドログラムで描画した。

最新の行動のクラスタの場所をわかりやすくするためにデンドログラムで「最新」のラベルを赤い色で付けた。

## 3.5 センサデータフロー

12/16

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

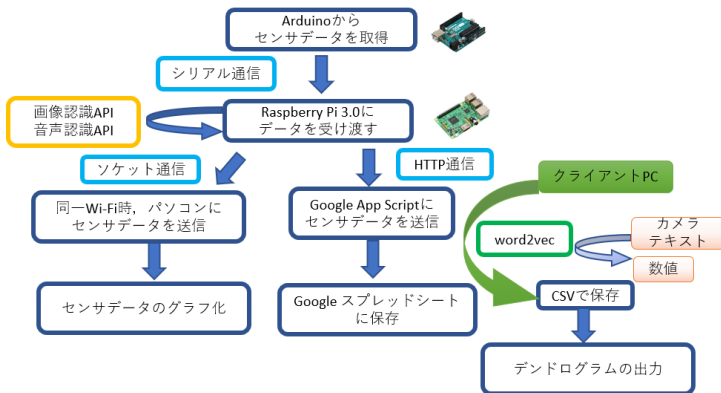


Figure 8: センサデータフロー

## 4.1 数値実験

13/16

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

### 実験設定

平日の 13 時から 18 時の 3 日のあいだにデータを収集した。データの収集中、行動に制限はない。ただし、デンドログラムにつけるラベルは多いほうがクラスターのまとまりの意味が分かるため、マイクによる音声入力 is 定期的に行っていく必要がある。今回の実験でライフログデータの取得中に行った動作として、パソコンの作業や外出したり、外食したりなどの行動をとった。

## 4.2 数値実験結果

14/16

1. はじめに
2. ライフログと各種センサ
3. 提案手法
4. 数値実験ならびに考察
5. 今後の課題

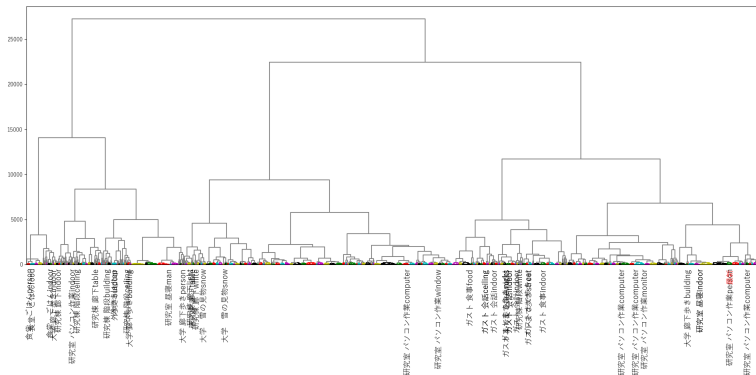


Figure 9: デンドログラムの結果

## 4.2 数値実験結果

15/16

### 実験結果

図 10 のデンドログラムの結果から、同じラベルは近いクラスターに分類された。

ただし、一部の行動が同一の行動であるはずなのに違うクラスターに分かれてしまう問題や、似た行動をしているのにそれらのクラスター分析でユークリッド距離が大きく離れてしまうところがあった。

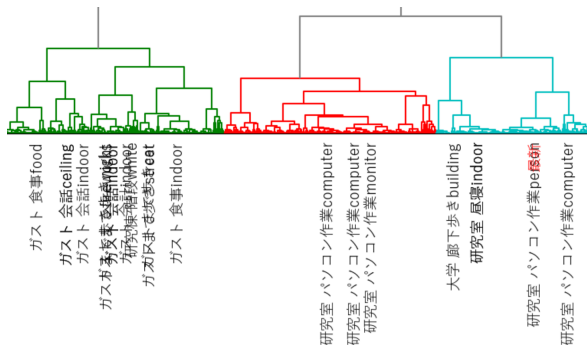


Figure 10: デンドログラムの拡大図

## 5. 今後の課題

16/16

### 今後の課題

- 同一の行動であるはずなのに違うクラスタに分かれてしまう問題と似た行動であるのにユークリッド距離が大きく離れたクラスターに分かれる問題がある。  
結果で得られるデンドログラムの縦の線をうまく活用することでそれぞれのクラスターを明確に区別できるようにする。
- Google の外部サーバーを使って実験を行っていたため、もしも Google のサーバーが落ちてしまったときアクセスできなくなってしまう問題と無料のサービス内で実験を行うにはいくつか制限がある問題がある。  
研究室のサーバーを使うことで、外部のサーバーに依存することなく制限もない状態にする。