

January 10, 2020

クラウド経由のデータ取得

沼田 賢一

富山県立大学 情報基盤工学講座

1. はじめに
2. セクション名
3. おわりに

January 10, 2020

はじめに
提案手法
クラウド
おわりに

本研究の背景

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う。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。

目的

従来の提案手法による行動パターン解析における分類精度の向上を図る。

提案手法

3/13

分類精度向上のための提案手法

- ① 入力信号によるラベル付け
- ② テキストデータのベクトル化

はじめに
提案手法
クラウド
おわりに

3/13

入力信号によるラベル付け

4/13

ライ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ライ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。

4/13

テキストデータのベクトル化

5/13

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。

5/13

目的

6/13

はじめに
提案手法
クラウド
おわりに

今回の目的は、自分の PC とラズパイのデータの受け渡しをソケット通信で行っていたものを、クラウドを経由するようにすることである。

今までの問題点として、すでにあるプログラムでは自分の PC とラズパイが同一 wifi 内でしかデータの受け渡しができていなかった。実際に、データを集めていくことを想定するとラズパイと PC が常に同じ wifi の中でいる事はできない。そこで、IP アドレスの固定だったりをしなくてはいけないが難易度が高いので一旦クラウドに保存して自分の PC からクラウドに定期的にアクセスしてデータの受け渡しをできるようにした。

6/13

クラウドといっても、たくさんあるのが無料でできる範囲で自分が使ったことのある google spreadsheet を使った。

はじめに

提案手法

クラウド

おわりに



データの POST

8/13

ラズパイからクラウドにデータを送るときに通信するやり方として、HTTP 通信の POST を使った。

また、HTTP 通信をするために `request` モジュールを用いた。データは、一度ここで JSON に変換して送信するようにした。

クラウドでの処理

9/13

クラウドでの処理として、スプレッドシートに紐付けされた Google App Script で POST された JSON データの受け取りから spreadsheet への保存を Javascript で書く。

はじめに
提案手法
クラウド
おわりに

9/13

クラウドへのアクセス

10/13

自分の PC からスプレッドシートにアクセスするために Google Sheets API を使って、取得するようにした。また、スプレッドシートへの認証は oauth で認可できるようにしている。

はじめに
提案手法
クラウド
おわりに

10/13

自分の PC での処理

11/13

クラウドでデータの取得とすでにあるデータは除いて csv 追加して保存し、その csv からデンドログラムを表示できるようにした。

11/13

はじめに

提案手法

クラウド

おわりに

クラウドの制約

12/13

まず、google app script の制約としてトリガーに対しての一日の合計実行時間が一日に 1 時間であることである。そのため、センサのデータを一つづつ送ると一日の制約にすぐ引っかかってしまった。
(一つづつだと受け取り保存までに 0.4 0.5 秒)

そのため、ラズパイからクラウドに送るデータはまとめて送り実行時間をへらす必要があった。計算では 2 分ごとがギリギリであるが余裕をもたせて 2 分半ごとにデータを送るようにした。

さらに、Google Sheets の外部からのアクセスに対しても制約があり 100 リクエスト / 100 秒・ユーザーというものもあるため、調整した。

12/13

まとめ

- ① 定期的にデンドログラムが更新できるようにした。
- ② もともとはセンサのデータはまとめて送ってなかつたので、2次元の配列にしたのでなかなかラズパイから自分のPCまでデータを移動するのになかなか扱いづらかった。
- ③ データのマイクラベルの時間とセンサの時間を対応するようになしたいがまだうまく行ってない。