

環境認識ライ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

富山県立大学 工学部 電子・情報工学科
1415048 福嶋瑞希

指導教員：奥原浩之

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、アプリケーションを用いて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た活動の記録は、ライログ [1] と呼ばれ、このライログログデータから行動のパターンを解析することで、個人に対しては、というログから健康管理や生活の見直し、社会に対しては、広告などのマーケティングに生かすことができる。

2 目的

このように個人や社会に生かすことを期待されているライログだが、現在ライログには大きく分けて二つの問題がある。

一つ目はライログの個人情報問題である。現在のライログアプリケーションは、GPS をライログとして取得、解析するアプリケーションが多く存在する。しかし GPS はライログデータのなかでも精密な個人情報が含まれるため、情報漏えいのリスクに対する警戒心が強いのが現状である。

二つ目は、ライログの煩雑問題である。ライログアプリケーションの中には位置情報や行動などのデータを自動的に記録するアプリケーションと、手動で記録しなければならないアプリケーションがある。手動で記録するということは、ライログデータを自身で取捨選択できるため個人情報保護につながるが、手間がかかり、めんどうであるという問題となる。

本研究ではこの問題に対し、GPS を使用せず、個人情報保護に着目し、スマートグラス EPSON MOVERIOBT-300 を用いて自動的に取得できる手間がかからないライログシステムを開発する。また、取得できるデータから行動パターンの類似性やイベント性を検出・考察することを目的とする。

3 行動識別

本研究で開発するシステムは、スマートグラスと画像認識を用いた視界情報取得アプリケーションを用いたデータ取得部と、多変量解析によるデータ解析を用いた行動識別部で構成される。データ取得部であるアプリケーションは、まずスマートグラスのカメラ機能を起動し、カメラ画像を取得、画像認識 API に送信すると、その画像に対するタグ情報やキャプション情報を取得できる。このテキストデータを保存して、約一分周期でテキストデータを取得するアプリケーションとなっている。このとき、省電力化としてカメラ機能を終了させているため、3 時間程度連続で使用可能となっている。

行動識別部は多変量解析を用いた結果に解釈、考察を行い行動の特徴や類似性を考察する。このとき、自己組織化マップのプロット間に線を引くなど工夫を行っている。

4 提案手法

本研究で開発するシステムは、スマートグラスと画像認識を用いた視界情報取得アプリケーションを用いたデータ取得部と、多変量解析によるデータ解析を用いた行動識別部で構成される。データ取得部であるアプリケーションは、まずスマートグラスのカメラ機能を起動し、カメラ画像を取得、画像認識 API に送信すると、その画像に対するタグ情報やキャプション情報を取得できる。このテキストデータを保存して、約一分周期でテキストデータを取得するアプリケーションとなっている。このとき、省電力化としてカメラ機能を終了させているため、3 時間程度連続で使用可能となっている。

行動識別部は多変量解析を用いた結果に解釈、考察を行い行動の特徴や類似性を考察する。このとき、自己組織化マップのプロット間に線を引くなど工夫を行っている。

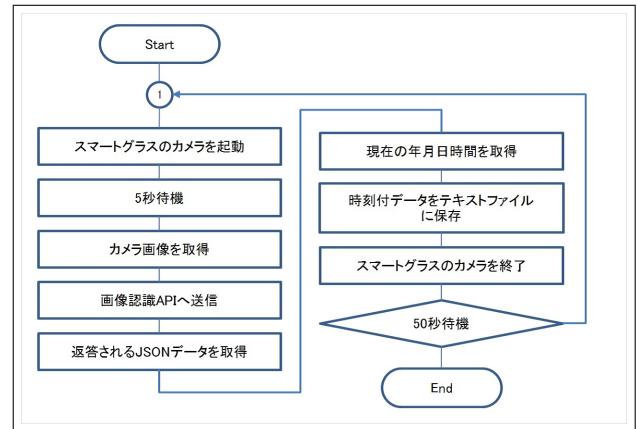


図 1: アプリケーションのフローチャート

5 数値実験と考察

提案手法を用いて数値実験を行った。アプリケーションを使用して、三時間分データを取得した。使用するデータは毎分の画像データから得られキャプション情報一つと、タグ情報5つである。このときデータ1はPC作業と、外出などの行動から構成され、データ2はノートPCでの作業、食事などから構成される。このデータ1とデータ2は別日であり、場所も異なっている。この二つのデータをあわせたcsvファイルをデータ3とした。

クラスター分析

まずクラスター分析を行う。クラスター分析はクラスターごとに集まっている抽出語からデータ3の傾向や特徴を知ることができる。データ1はピンクのクラスターにはいり、同じクラスターには computer や desktop という単語が含まれることからデータ1はデスクトップPCが視界に入る作業を行っていることがわかる。また、データ2は、laptop や screen shot という単語が多く含まれ、ノートPC作業が多いと考えられる。赤のクラスター、青のクラスターとともに外出を表すものが含まれている。黄色のクラスターはキッチンや食事、家具を表す単語が含まれ、この三つの行動は同じくらいの時間だと考えられる。

多次元尺度法

抽出語間の関連性や類似性の強さがプロット間の距離にあらわされる [2]。PC作業に関する単語から構成されるクラスター01とクラスター02が近く、それ以外のクラスターが一定の距離を保つて離れていることから、中心的な作業とそうではない作業があることがわかる。また、データ1とデータ2は近くのクラスター同士にプロットされていることから、類似する行動がPC作業という行動であることがわかる。また、データ1は外出という行動から近いが、食事を表す単語はデータ2の方が近い。このことより、データ1、データ2は類似する行動としてPC作業、イベント性がある行動として食事や外出であると考えられる。

対応分析

近くに位置しているものは相対的に関連が強く、特徴のない語が原点付近に密集することが多い分析手法である。データ1、データ2共に同じくらい出現している抽出語、つまり特徴的ではない抽出語として indoor や computer が出現している。このことから両方のデータは室内での作業、PC作業が含まれているのではないかと考察できる。データ1からみて、データ2に含まれる table 等は関係性が近いが、food 等は関係性がないため特徴的であるように出力されている。同じようにデータ2からみて、データ1に含まれる snow 等は特徴的な語となっている。この比較より、データ1とデータ2は indoor や computer が含まれる行動、室内行動やPC作業が類似性のある行動であることがわかる。また、データ1は外出という行動がイベント性があり、データ2は食事や外出であることがわかる。

タ 2 は食事という行動がイベント性のある行動となっていることがわかる。

共起ネットワーク

線がつながっている語同士は関係性があることをしめす分析手法である[3]。データ 1 とデータ 2 はともに table, desk, computer, indoor という抽出語と共に関係があり、両方とも強い共起関係をもつのは PC 作業を表す抽出語であり、類似性のある行動であることが確認できる。また、同じ PC 作業が示されていることがわかるが、データ 1 には desktop, データ 2 には laptop が含まれることから同じ PC 作業でも使っている PC が違うことがわかる。

自己組織化マップ

似たテキストを近くにプロットしていく手法である。データ 1 は赤色の線分、データ 2 は青色の線分で示す。これより、データ 1 とデータ 2 の時系列は類似性が低いことがわかる。理由として、同じ PC 作業であってもデスクトップ PC とノート PC という別の PC を使用した作業であるため同じ行動の中でも視界に写る物体が違いから行動が区別されているからであると考える。また、赤い線と青い線が両方ともつながっている時間、黄色で部分はコンピュータースクリーンに近づいて作業を行っていることや person という単語が入るテキストが含まれていた。これは PC 画面に映った人の画像を認識していると考える。

6 おわりに

最後におわりにして結論と今後の課題。結論として、個人情報保護に着目したライフログデータ取得アプリケーションの開発ができ、多変量解析を用いることでライフログの可視化を行い行動パターンの類似性やイベント性を視覚的に検出するという目標は達成できた。特に、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ため、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るために、GPS を使用しなくても、ライフログデータに位置情報を付加できると考えられる。よって、個人情報保護に着目し取得したライフログデータから行動パターンの類似性やイベント性を検出できたと考える。

今後の課題として、開発したアプリケーションは自動的にライフログデータを取得する点が利点として挙げられるが、一方で客観的なライフログデータしか取得できないという弱点もある。ユーザーが興味を持った瞬間や、データを取得したい瞬間のライフログデータは現状のアプリケーションには含まれていないためである。この弱点に対し、ユーザーが取得したいタイミングでライフログデータを取得する方法をアプリケーションに組み込む必要がある。また、アプリケーションの稼働時間を伸ばし、より多くのデータを収集できるようにすることを今後の課題とする。

参考文献

- [1] 新保史生, “ライフログの定義と法的責任 個人の行動履歴を営利目的で利用することの妥当性”, 情報管理, Vol. 53, No. 6, pp. 295–310, 2010.
- [2] 斎藤堯幸, “多次元尺度構成法”, 計測と制御, Vol. 22, No. 1, pp. 126–131, 1983.
- [3] 田中京子, “KH Coder と R を用いたネットワーク分析”, 久留米大学コンピュータジャーナル, Vol. 28, pp. 37–52, 2014.