

卒業論文

環境認識ライフ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

Similarity and event detection by behavior pattern analysis from
environment recognition life log

富山県立大学 電子・情報工学科

1415048 福嶋 瑞希

指導教員 奥原 浩之 教授

平成30年2月6日

記号一覧

以下に本論文において用いられる用語と記号の対応表を示す.

用語	記号
クラスター	X, Y
クラスター内での重心とサンプルとの距離の 2 乗和	$L(X), L(Y)$
クラスターの重心とクラスター内の各サンプルとの距離の 2 乗和	$L(X \cup Y)$
入力データベクトル	x
競合層のニューロンの番号	i
参照ベクトル	m_i
勝者ニューロン	c
勝者ニューロンとの距離によりガウス関数で減衰する係数	h_{ci}
i 番目のニューロンの競合層上での位置	r_i
勝者ニューロンの競合層上での位置	r_c
学習回数	t
学習率係数	$\alpha(t)$
学習半径	$\sigma^2(t)$

目次

記号一覧	1
第1章 序論	5
§ 1.1 本研究の概要	5
§ 1.2 本研究の目的	5
§ 1.3 本論文のあらまし	6
第2章 ライフログとスマートグラス	7
§ 2.1 現状のライフログ	7
§ 2.2 スマートグラス	8
§ 2.3 画像認識 API	10
第3章 行動識別	13
§ 3.1 行動識別	13
§ 3.2 行動識別のための分析手法	14
§ 3.3 類似性・イベント性	19
第4章 提案手法	23
§ 4.1 開発システムの概要	23
§ 4.2 省電力化	24
§ 4.3 周期性の検出	25
第5章 シミュレーション結果ならびに考察	27
第6章 結論と今後の課題	35
謝辞	36
参考文献	37
付録	40
A. 1 ライフログデータ取得アプリケーションのソースコード	40
A. 2 時系列 SOM を作成するソースコード	42

序論

§ 1.1 本研究の概要

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。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ライフログデータは、個人の生活に生かしたり、社会に生かしたりできると考えられている。

スマートフォンやウェアラブルデバイスの全地球測位システム (Global Positioning System, Global Positioning Satellite : GPS) や撮影情報をライフログとして取得、解析するアプリケーションが多く存在し、受容性の高いライフログのための研究が行われている [1]。しかし GPS や撮影情報はライフログデータを取得する本人だけでなく、第三者に対する個人情報が含まれるため、不安や嫌悪感が大きく、情報漏えいへのリスクに対する警戒心が強いのが現状であり、技術面とは異なる課題となっている [2]。また、手動でライフログデータを取得するアプリケーションも多く、未だライフログの受容性は改善の余地がある。

個人情報に対する心理的不安、ライフログデータを取得するという物理的負担が少ないライフログは多くの人に広く受け入れられると考える。ライフログ自体が多くの人に広く受け入れられることで、取得するデータ量を増やすことができるため、より個人や社会に生かすことができると考えられる。したがって、ライフログの在り方は改善すべきであると考えられる。

§ 1.2 本研究の目的

本研究は、多くの人に広く受け入れられるライフログとして、個人情報保護に着目し、手間がかからず自動的にライフログデータの取得を行い、取得したデータから類似性やイベント性を考察できることを目的とする。この目的のため、スマートグラスと画像認識を用いたリアルタイム視界情報取得アプリケーションを提案する。このアプリケーションを使用したビッグデータ構築、データ解析を行い、行動パターンの類似性・イベント検出を行

う。また、このアプリケーションの開発には画像認識 API を使用し、デバイスは EPSON MOVERIOBT-300 を使用する。データの解析は、自己組織化マップ (Self-organizing maps: SOM)、階層的クラスター分析、多次元尺度構成法 (Multi Dimensional Scaling: MDS)、対応分析、共起ネットワークを行い、読み取り・比較を行うことでライフログデータの類似性やイベント性を考察する。

§ 1.3 本論文のあらまし

本論文は次のように構成される。

第 1 章：本章 第 1 章では、本研究の概要と目的について説明した。

第 2 章 第 2 章では、現状のライフログ・ライフログアプリケーションの問題や特徴について説明する。また、ウェアラブルデバイスであるスマートグラスの種類や本研究で使用するスマートグラスについての説明、視界情報を取得するための画像認識 API について述べる。

第 3 章 第 3 章では、行動識別についての研究や、行動識別のための分析手法について説明する。また、分析から考えられる類似性やイベント性に説明する。

第 4 章 第 4 章では、本研究の提案手法として開発した視界情報取得アプリケーションについて述べる。また、このアプリケーションを開発する上で、省電力化を行うことについて説明する。開発したアプリケーションを用いて取得したライフログデータの SOM をよりわかりやすくするための工夫についても述べる。

第 5 章 第 5 章では、開発したアプリケーションで取得したライフログデータから多変量解析を行ったうえでの行動パターンの類似性・イベント検出について考察を述べる。

第 6 章 第 6 章では、まとめと今後の課題を述べる。

ライフログとスマートグラス

§ 2.1 現状のライフログ

ライフログ (lifelog) とは、人間の活動 (life) の記録 (log) であり、センサーなどで個人の活動に関するログを取得する行為が、元来のライフログの語源と考えられている。本研究では、この行為をライフログとし、個人の行動履歴に基づいて生み出されるビッグデータのことをライフログデータと呼ぶこととする。また、ライフログに関して、長時間の記録や膨大なデータが必要という定義はない。

ライフログデータを取得・活用できるアプリケーションとして、ソニーモバイルの Xperia 専用アプリ「Lifelog¹」がある (図 2.1 参照)。このアプリケーションはスマートウェア「SmartBand 2²」 (図 2.2 参照) と連携することで歩数や心拍数等のライフログデータを取得し、ユーザ自身が健康管理に生かすことができる。また、自動で位置情報をマップにマッピングできる「マッピング - GPS ログまとめて全部記録³」 (図 2.3 参照) や、手動でマッピングできる「Swarm⁴」 (図 2.4 参照) というアプリケーションがある。このアプリケーションは行動の記録を取ることができるため、日々の生活や旅行の記録として使用できる。上記のアプリケーションは GPS のアクセス許可が必要であり、上記以外のライフログアプリケーションも GPS を必要とすることが多い。

ライフログに関する既存研究として、スマートフォンから得られる位置情報履歴や写真撮影履歴、ツイートを使用したライフログデータから行動特徴抽出・イベント検出を行う研究が行われている [3] [4]。取得したライフログデータの解析を行うことで、ユーザー自身の健康管理や学習 [5] に生かすだけでなく、ビジネスとしてターゲティング広告に生かすこともできる。ライフログは、様々な視点からライフログデータの比較を行うことで、個人や社会に利用できるという価値があると考えられる。

しかい、現状のライフログには、大きくわけて二つの問題があると考えられる。一つ目

¹<http://www.sonymobile.co.jp/myxperia/app/lifelog/>

²<http://www.sonymobile.co.jp/product/smartproducts/swr12/>

³<https://play.google.com/store/apps/details?id=org.liteapp.mat2>

⁴<https://play.google.com/store/apps/details?id=com.foursquare.robin>

はライフログの個人情報問題，二つ目はライフログの煩雑問題であると考えられる，

ライフログの個人情報問題

ライフログデータとして主に用いられることが多いのは，GPS であると考えられる．GPS を用いることで，正確な位置情報を取得することができるため，いつ，どこに，どれくらいいるのかという情報をライフログデータに含むことができる．また，同時にツイートやその場で撮影した写真を取得することで，どんな行動を行っているか推測することができる．正確なライフログデータを取得できる反面，GPS の情報がネット上でどのように扱われているかユーザーは把握できず，一度情報が漏えいしてしまうと個人が特定されてしまうというリスクがある [6]．このような，GPS の含まれたライフログデータという個人の活動に関するログは，個人を特定することが安易であるため，個人情報の取り扱いに伴う義務が生じプライバシーの侵害という問題を引き起こす [7]．

ライフログの煩雑問題

ライフログアプリケーションの中には，意識的にライフログデータを取得しなければならないアプリケーションが存在する．このようなアプリケーションはユーザーが自ら位置情報をマッピングしたり，ライフログのために食事風景の写真をとることを意識しなくてはならない [8]．ユーザーの主観的なライフログデータを取得できるが，ライフログデータを取得するのに手間がかかってしまうという問題を引き起こす．

また，ライフログの個人情報問題に関して，株式会社NTTデータ経営研究所が2016年に10代から60代の男女1059人を対象として実施した「パーソナルデータに関する一般消費者の意識調査 [9]」という調査がある．この調査において，「企業のマーケティング等の利用目的にて，パーソナルデータを企業に提供しても良いと思うデータの条件」において，金銭や商品を受け取ることができたり，個人が特定できない状態でも，どのような条件であっても位置情報は提供したくないという人が66.2%であり，過半数以上を占めていることがわかっている．

ライフログの個人情報問題，煩雑問題という二つの問題に対し，ライフログデータを収集する上で重要であることは，可能な限り不安要素を取り除くことと，手間をかけず無意識でライフログデータを残すことであると考えられる．GPSを使用せず，自動でライフログデータを残すことを可能にすることにより，誰でもプライバシー侵害の不安や負担のないライフログを可能にする．

§ 2.2 スマートグラス

近年，スマートフォンやタブレットなどのスマートデバイスが急速に普及している．その次のデバイスとして期待されているものがウェアラブルデバイスである．ウェアラブルデバイスとは，体に装着して利用するコンピュータデバイスの総称であり，スマートグラスはメガネ型のウェアラブルデバイスである．代表的なものとして，Glass Enterprise Edition⁵」

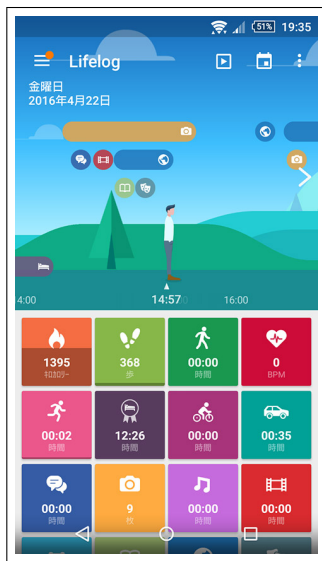


図 2.1: Lifelog



図 2.2: SmartBand 2

(図 2.5 参照), SmartEyeglass⁶ (図 2.6 参照) などが挙げられる。

このようなスマートグラスは把持の必要がなく、常に目の前に仮想画面を表示可能である [10] ため、両手を常に開けておくことが可能である。手をあまり使うことなく欲しい情報を提示することができ、また、外部から見ているものを知られることなく情報を活用できる [11]。さらに、ユーザーが見ている実際の景色に必要な情報を重ねて表示することができるため、拡張現実技術との親和性も高い。

このように、スマートグラスを使用すると、ユーザーがあまり意識する事なく、常時画像の取得を行える。この利点を生かし、ユーザーに負担をかけることなくライフログデータを取得できる。

本研究では、スマートグラスとしてセイコーエプソン社のシースルーモバイルビューアー MOVERIO BT-300 を使用する (図 2.7)。使用する理由として、スマートグラスの中でも比較的安価であり、Android アプリケーションを作成して動作できるためである。MOVERIO はユーザーが見ている現実空間に対してコンピュータが生成した仮想オブジェクトを重畳表示するというシースルー表示を行う。シースルー表示の実現にはビデオシースルー方式と光学シースルー方式の 2 通りがあり、MOVERIO は光学シースルー方式を使用することが可能である [10]。ビデオシースルー方式は、カメラ画像とコンピューターグラフィックス (CG) を合成した画像を表示する方式である。一方、光学シースルー方式は肉眼の視界に対して CG を重畳する方式であるため、視界が広く、移動中の使用や現実の物体を用いた作業時の使用に適している。

⁵<https://www.x.company/glass/>

⁶<https://developer.sony.com/ja/develop/smarteyeglass-sed-e1/>

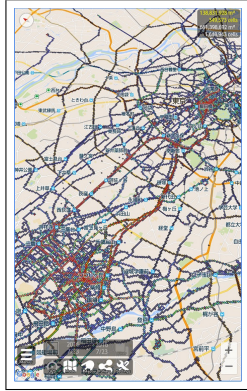


図 2.3: マッピング-GPS ログまとめて全部記録

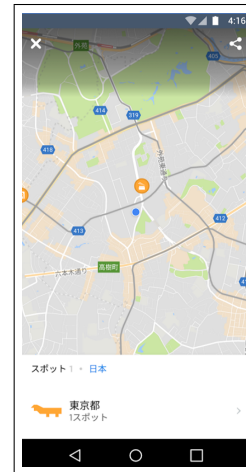


図 2.4: Swarm



図 2.5: Glass Enterprise Edition



図 2.6: SmartEyeglass

§ 2.3 画像認識 API

画像認識技術とは、コンピュータに画像を理解をさせる技術である。画像内のピクセル信号のパターンから意味を抽出するパターン認識により、人間の視覚機能をコンピュータに処理させることができる。画像認識技術を用いることで、画像が持つ情報をテキストで取得することができる。代表的な画像認識 API に、Google Cloud Vision API と、Computer Vision API, IBM Bluemix Alchemy API という三種類がある。画像認識 API を使用することで、個人だけでは取得が難しい大量のデータを利用することができるため、スマートフォンやウェアラブルデバイスで取得したカメラ画像を認識するアプリケーションの開発が可能となる。



図 2.7: MOVERIO BT-300

Google Cloud Vision API

2016 年に一般公開された画像認識クラウドサービス Google Cloud Vision API⁷は、写真の被写体を機械判定し、ラベリングする機能をもっている。Google Cloud Vision API を用いて個々の写真の情報（ラベル）を取得できるが、撮影内容が不明瞭のときは、1 つも得られないこともあり [12]、同じ単語を複数個返して来る場合もある。初年度無料である。

Computer Vision API

Microsoft 社が提供している API の一つである Computer Vision API⁸は、写真画像の被写体を機械判読し、ラベリングや画像内のテキストの判読など多様な機能を有している。ラベリングだけでも tags や captions という機能を有する。tags は、画像内の要素を、2,000 以上の認識要素、生物、風景などに基づいて、タグ情報を算出する [13]。captions は文章で人間が読める言語として要約を表示する。初年度無料である。

IBM Bluemix Alchemy API

Watson が提供している API の一つである IBM Bluemix Alchemy API⁹は、画像に対してタグ付けを行うことができる。キャプションは出力できないが、分類結果の階層構造に強く、食べ物の分類などに特化している。Lite コースは無料である

⁷<https://cloud.google.com/vision/>

⁸<https://azure.microsoft.com/ja-jp/services/cognitive-services/computer-vision/>

⁹<https://www.ibm.com/watson/jp-ja/developercloud/visual-recognition.html>

行動識別

§ 3.1 行動識別

携帯電話やウェアラブルデバイスを用いて、ユーザーが今何を行っているかという行動をライフログデータとして取得し、取得したライフログデータの解析から行動を認識することを行動認識や行動識別という。本研究では、行動識別と呼ぶことにする。

既存研究には、ライフログデータを取得するため、携帯電話の加速度センサやGPSを用いて走行や歩行しているなどの行動識別を行う研究 [14] や、ウェアラブルデバイスの加速度センサやGPSを利用する事で人の行うさまざまな行動を取得し、行動識別を行う [15] 研究がある。しかし、本を読んでいることであったり、料理をしていることなどの細かい動作をライフログデータとして取得することは難しい。机上に設置した KinectTM(図 3.1 参照)を用いて机上の細かい動作を認識する研究が行われているが、机上に限っているため屋外や机上以外の行動は認識できない [16]。なお、KinectTM は 2017 年 10 月 25 日に生産終了が公表されている。

本研究ではGPSやKinectTMを使用せず、細かい行動をライフログデータとして取得する手法として、ユーザーの視界情報を取得する。視界情報を取得することで、視界上にある物体からどのような作業を行っているのか、視界の風景から室内か室外にいることなどを判断できるため細かい行動をライフログデータとして取得することが可能となる。しかし、ユーザーの視界情報を取得すると、GPSを使用しなくても、どこで何をしているのか、誰と会っているのかなど必要以上のライフログデータを取得してしまう。この結果、ユーザーやユーザーの視界にいる第三者のプライバシーを侵害してしまう。さらに、視界情報を画像や動画で蓄積するとデータ量が多くなるという問題が生じる。この問題に対し、ウェアラブルデバイスのカメラ画像を取得し、減色処理を施しデータの蓄積・解析を行う研究が行われている [17]。本研究ではMOVERIOのカメラ画像を取得し、画像認識によりカメラ画像をテキストに変換することで、データ量を削減、かつプライバシーに配慮したライフログデータの取得を検討する。



図 3.1: Kinect™

§ 3.2 行動識別のための分析手法

本研究ではいくつかの解析手法を用いて、一定時間内の取得データを視覚的に表し、行動識別を行う。そのために、多変量解析である SOM、階層的クラスター分析、MDS、対応分析、共起ネットワークを用いてテキストデータの可視化を行う。

多変量解析を行うツールとして KH Coder¹⁰を使用する。KH Coder とは、テキスト型データの計量的な内容分析、もしくはテキストマイニングのためのフリーソフトウェアである [18]。無償でウェブサイトから入手でき、すべての機能をマウス操作で利用できる。また、どんな言葉が多く出現していたのかを頻度表から見ることができたり、SOM、共起ネットワークなどの多変量解析を行ったりできる [19]。KH Coder を用いて行われた研究としては、アンケートの自由回答項目・新聞記事・インタビューデータなどさまざまなデータを分析した事例がある [20]。本研究では Version 3.Alpha.11 を使用する。また本章ではチュートリアル用に KH Coder から提供される「坊ちゃん」英語版テキストデータを用いて解析を行う（図 3.2 参照）。この「坊ちゃん」英語版テキストデータは KH Coder をダウンロードする際に同時に取得することが可能であり、KH Coder のホームページよりダウンロード¹¹後 C:\khcoder3\tutorial.en にテキストファイルとして保存されている。

KH Coder をダウンロードする際に取得できる KH Coder3 リファレンス・マニュアルによると、KH Coder を使用した多変量解析には、まず対象のテキストファイル（もしくはエクセルファイル）を読み込むことから始める。読み込むテキストファイルに事前に手動で h1 タグや h2 タグという見出しタグを設定することで、見出しごとの解析も可能である。この時 Stop words として Be 動詞のような一般的な語を指定しておく、分析対象から外すことができる（図 3.3 参照）。この Stop words 一覧テキストファイルは、KH Coder をダウンロードする際に同時に取得することが可能であり、解析を行う際に必要でない単語を自由に追加できる。

次に、前処理として、POS Tagger¹²を使用して自然言語処理を行う（表 3.1）。前処理を行うことで、多変量解析に使用する「各文書に、それぞれの語が何度出現していたのかという集計表」である「文書×抽出語」表 [21] (表 3.2 参照) を出力することができる。h1 から h5 というのは見出し番号であり、h1 タグや h2 タグが存在すると 1 増加する。dan は段落番

¹⁰<http://khc.sourceforge.net/>

¹¹<http://khc.sourceforge.net/dl3.html>

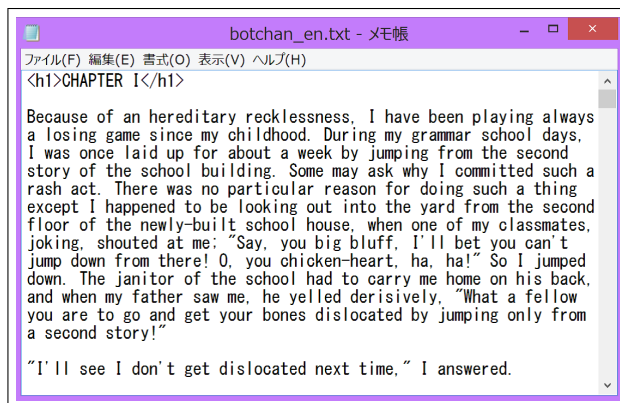


図 3.2: 「坊ちゃん」英語版テキストデータの一部

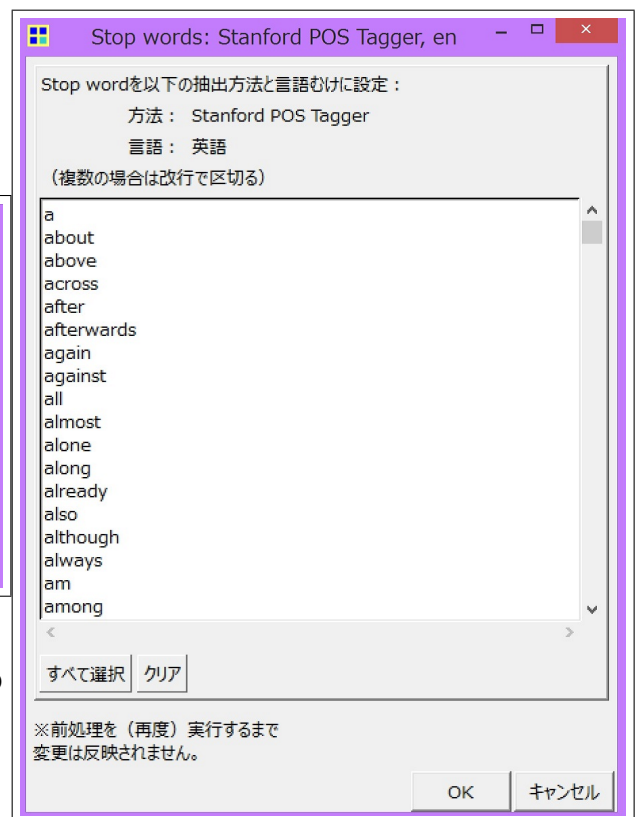


図 3.3: Stop words の一部

号で, bun は文番号である. id は文書の通し番号で, リセットされることは無い. length_c は文書の長さを文字数で表し, length_w は文書の長さを語数で表したものである.

まず, ライフログデータの内容解析のため, 階層的クラスター分析, MDS, 対応分析, 共起ネットワークを行う. この時テキストデータは, 抽出語の中でも, 50 回以上出現する 35 語の抽出語を用いる. 理由として, 出力される抽出語が多すぎると解析結果の読み取りが難しくなるためである.

階層的クラスター分析は, 抽出語の最も似ている組み合わせから順番にクラスターにしていく方法であり, デンドログラムを表示する [22]. 指定されたクラスター数に全体を分割し, その結果を色分けによって表示する. なお, KH Coder では, デフォルトの Auto では, 抽出語数の平方根を四捨五入したものをを用いている [23]. 1つのクラスターには関連性が高い抽出語が集まっているため, クラスターごとに集まっている抽出語を調べることでテキストデータ全体における文書の傾向や特徴を知ることができる. 階層的クラスター分析の作成方法は, まず抽出語として A,B,C,D があったとする. この時抽出語の中で最も距離の近い組み合わせを A と B とし, A と B をくくり, 2 点の代表点を求める. 次に, AB の重心, C, D の 3 点で, 最も距離の近い組み合わせを見つける. このとき C と D が最も近いとすると, C と D をくくる. このように繰り返していくことで, デンドログラムを作成する [22] [24].

KH Coder3 リファレンス・マニュアルによると, クラスター間の距離測定方法として,

¹²<https://nlp.stanford.edu/software/tagger.html>

表 3.1: 「坊ちゃん」データを用いて KH Coder で出力した抽出語の一部

Noun		ProperNoun	
school	130	Red	171
room	119	Shirt	163
teacher	119	Porcupine	128
house	108	Clown	85
time	95	Kiyo	73
fellow	88	Tokyo	47
day	84	Hubbard	46
student	84	Squash	46
way	78	Badger	32
night	70	Madonna	28
head	65	Koga	23
face	64	Sir	21

表 3.2: 「坊ちゃん」データを用いて KH Coder で出力した「文書×抽出語」表の一部

h1	h2	h3	h4	h5	dan	id	length_c	length_w	school	room	teacher	house	time
1	0	0	0	0	1	1	650	177	4	0	0	1	0
1	0	0	0	0	2	2	49	16	0	0	0	0	1
1	0	0	0	0	3	3	203	51	0	0	0	0	0
1	0	0	0	0	4	4	43	15	0	0	0	0	0
1	0	0	0	0	5	5	207	58	0	0	0	0	0
1	0	0	0	0	6	6	577	147	1	0	0	1	0
1	0	0	0	0	7	7	853	216	0	0	0	0	0
1	0	0	0	0	8	8	800	193	0	0	0	1	1
1	0	0	0	0	9	9	155	42	0	0	0	0	0

KH Coder ではワード法を使用している．2つのクラスター X, Y を結合したと仮定したとき，それにより移動したクラスターの重心とクラスター内の各サンプルとの距離の2乗和 $L(X \cup Y)$ と，もともとの2つのクラスター内での重心とそれぞれのサンプルとの距離の2乗和 $L(X)$ ， $L(Y)$ の差が最小となるようなクラスターどうしを結合する手法である．ワード法は，計算量が多いが分類感度がいいため用いられることが多く，ワード法は一つのクラスターに抽出語が順に吸収され類似するクラスターが形成される鏡効果が起こりにくいという強みがある [25] [26]．

$$\Delta = L(X \cup Y) - L(X) - L(Y) \quad (3.1)$$

クラスター分析の結果を図 3.4 に示す．この時クラスター数を Auto にしたためクラスター数は6となる．併合標準（図 3.5 参照）からクラスター数が6であることは妥当だと考えられるためクラスター数は6とした．クラスター分析から，最も多く出現している say という単語があるクラスターに school という単語がある点から，学校で何かを話す場面が多いのではないかと推測できる．また，teacher と student が同じクラスターにある点からやはり

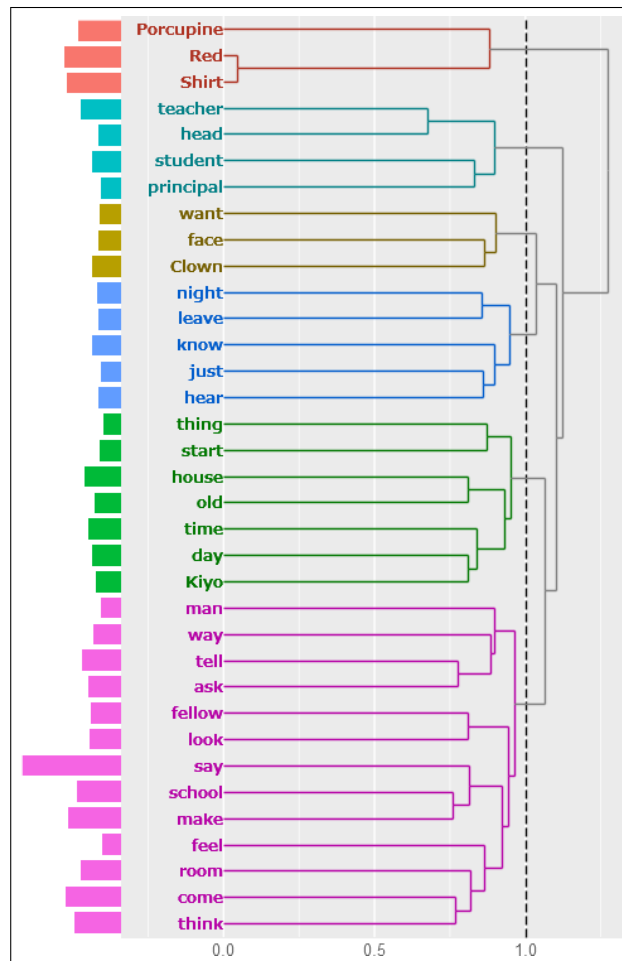


図 3.4: 「坊ちゃん」データから作成したクラスター分析

学校が重要ではないかと推測できる．クラスター内やクラスターどうしの比較からデータ内で重要な語の関係を推測できる．

MDS は，抽出語間の関連性や類似性の強さをマップ上の点と点の距離に置き換えて，相対的な関係性を視覚化する手法である [27]．KH Coder では MDS の中でも最も広く利用されてきた Kruskal の非計量 MDS を使用している [28]．また，語と語の関連を見るために Jaccard 係数を使用している．Jaccard 係数とは二文章間の類似度であり，「語 A を含む」かつ「語 B を含む」文書の数，「語 A を含む」または「語 B を含む」どちらかでも当てはまる文書の数で割った係数である [29]．MDS の結果は，相対的な位置関係だけを表わしているため軸の方向性に意味はない．

$$\text{Jaccard 係数} = \frac{\text{語 A と語 B を含む文書}}{\text{語 A もしくは語 B を含む文書}} \quad (3.2)$$

図 3.7 は坊ちゃんデータから出力した MDS である．クラスター分析を参考にし，クラスター数は 6 に設定した．この結果，目立つものはクラスター 01 であり，どのクラスターからも同じような距離であることから，データの中で中心的なクラスター・抽出語であることがわかる．クラスター 02 と 06 のように離れて配置されるクラスターもあり，関係性の

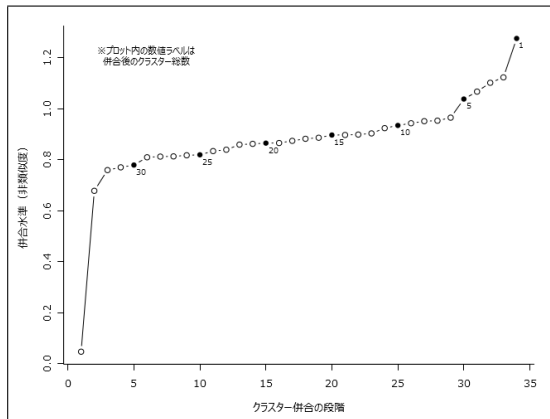


図 3.5: クラスター分析の併合標準

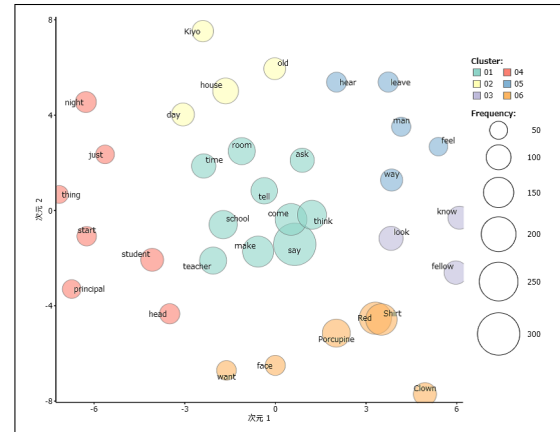


図 3.6: 「坊ちゃん」データから作成した MDS

低いクラスター・抽出語であることがわかる。

対応分析は、単純な 2 次元表や多重表の行と列間の対応する測定値を分析する探索的データ解析の手法であり、分析結果として、2 次元のマップが表示される [30]。このマップで近くに位置しているものは、相対的に関連が強いということを示し、遠くに位置しているものは関連が弱いということになる。また、対応分析では、これといって特徴のない語が原点付近に密集することが多い。

図 3.8 は坊ちゃんデータより出力した対応分析である。この対応分析から、think や say という行動は特徴的ではなく、データ全体によく出現することがわかる。一方で、Red Shirt や Clown は原点から遠く離れているため、特徴的であることがわかる。

共起ネットワークはある語が語られる状況の断面を多角的に把握するのに強力な解析手法であり、線がつながっている語が共起関係にあり、その繋がりにより注目する [31]。抽出語の中で出現パターンの似通ったものを線で結ぶネットワークであり、すなわち共起関係を線で表したネットワークである。MDS と異なっている点は、プロットされた位置ではなく線で結ばれているかどうかということに意味がある点である。

また、KH Coder3 リファレンス・マニュアルによると、KH Coder では、共起ネットワーク図の表し方として複数用意されており、それらの中から選択できる。種類は、中心性（媒介）、中心性（次数）、中心性（固有ベクトル）、サブグラフ検出（媒介）、サブグラフ検出（random walks）、サブグラフ検出（modularity）の 6 種類がある。この時の中心性が高い語とは、語がデータ中で重要な役割を果たしている可能性がある語である [32]。サブグラフ検出とは、比較的強くお互いに結びついている部分を自動的に検出してグループ分けを行い、その結果を色分けによって示す方法 [33] であり、抽出語同士の関係性が強い、つまり Jaccard 係数の値が高い抽出語の集まりである。本分析では、各抽出語の関係性を確認するため、KH Coder の出力する共起ネットワーク図の中から「サブグラフ検出（媒介）」を使用する [34]。

坊ちゃんデータを使用して出力した共起ネットワークが図 3.9 になる。この時、Jaccard 係数が 0.2 以上の共起関係を描画している。Jaccard 係数が小さいほど類似度が低いものも含まれ、大きいと類似度が大きいものしか描画されないため状況に応じて検討すべきである。共起ネットワークより、school は teacher や student と共起関係があり、make や say と

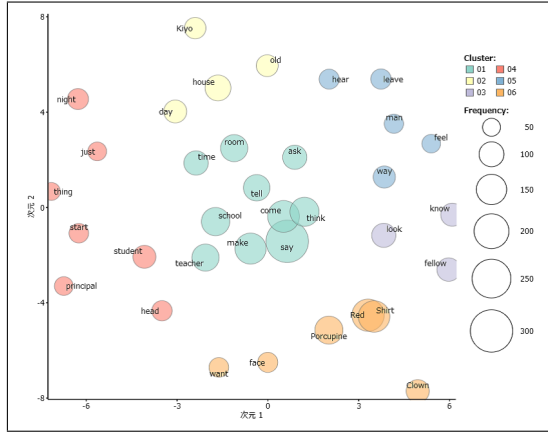


図 3.7: 「坊ちゃん」データから作成した MDS

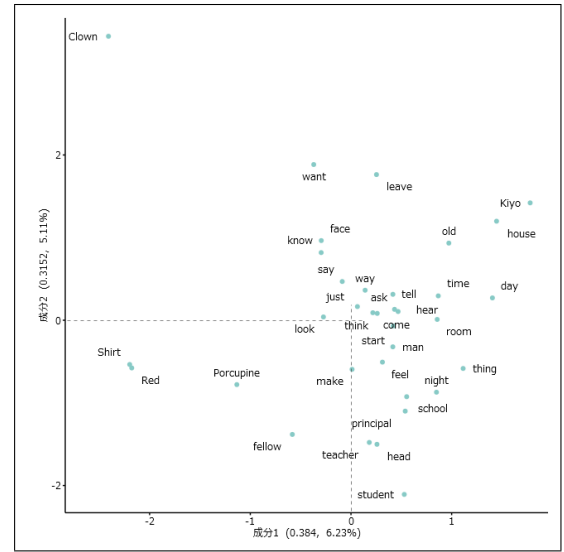


図 3.8: 「坊ちゃん」データから作成した対応分析

いう動詞とも関係性があることがわかる．room は come や think などの動詞と関係性があり，Red と Shirt は関係性があることが，視覚的に理解しやすくなっていると考えられる．

§ 3.3 類似性・イベント性

本研究では，ライフログデータの内容解析のため，階層的クラスター分析，MDS，対応分析，共起ネットワークを行った後，データの時系列を SOM を用いて解析を行う．SOM を用いて，テキストデータの時系列を可視化することでテキストデータの類似性を検出する．

SOM とは、ヘルシンキ大学のコホーネン教授により 1981 年頃に発表された，教師なし学習を行なうニューラルネットワークの代表例と言える解析手法である [35]．ニューラルネットワークとは、脳機能に見られるいくつかの特性を計算機上のシミュレーションによって表現することを目指した数学モデルである．つまり，人間が無意識にやっていることを機械にやらせるということである．

SOM は図 3.10 に示すように入力層と出力層の 2 つに分かれて競合学習を行う [36] [37]．図 3.10 には，入力層のニューロンが複数個あるが，各々のニューロンがそれぞれの次元に対応した出力を行っていると考ええる．入力データベクトルと呼ばれる入力層から出力層への入力を x と定義し，出力層のニューロンと入力層のそれぞれのニューロンとの結合強度は総称して参照ベクトルと呼ばれ， i を出力層のニューロンの番号とすると， m_i で表される．まず初めに， m_i の初期化を行い，入力データベクトル x を選び，入力データベクトルと各ニューロンの参照ベクトルとのユークリッド距離で出力層のニューロンを競合させる．勝者ニューロンを c とすると，式 3.3 で表される． $\arg \min f(a)$ は $f(a)$ を最小にする a の集合であり，下側に変数にとる値の範囲を書くことが多い．

$$c = \arg \min_i \{ ||x - m_i|| \} \quad (3.3)$$

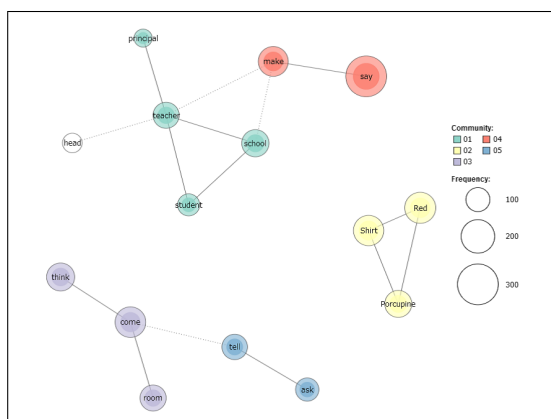


図 3.9: 「坊ちゃん」データから作成した共起ネットワーク

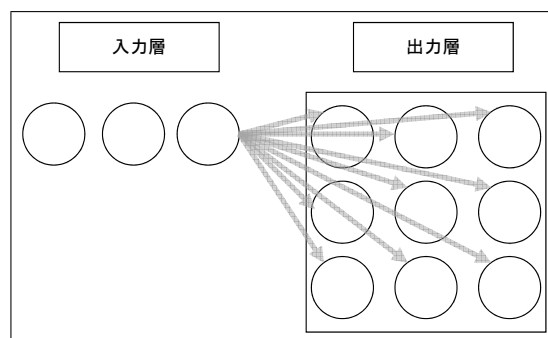


図 3.10: 入力層と出力層

次に、勝者ニューロンと勝者ニューロンに近いニューロンは自らの参照ベクトルと入力データベクトルを近づける学習を行うため、参照ベクトルを同様に更新させる。この時、 h_{ci} は勝者ニューロンとの距離によりガウス関数で減衰する係数である。

$$m_i(t+1) = m_i(t) + h_{ci}(t) \cdot \{x(t) - m_i(t)\} \quad (3.4)$$

$$h_{ci} = \alpha(t) \cdot \exp \frac{-\|r_c - r_i\|^2}{2\sigma^2(t)} \quad (3.5)$$

また、SOM 作成過程ではユークリッド距離を利用している。また、KH Coder3 リファレンス・マニュアルによると、文書の長さのばらつきに左右されない形で計算を行うために、文書中における語の出現回数をそのまま使うのではなく、1,000 語あたりの出現回数に調整したものを計算に使用している。

KH Coder3 リファレンス・マニュアルによると、KH Coder の SOM の学習は、大まかな順序づけを行う段階と、微調整を行う収束段階の 2 段階で行われる。KH Coder では、1 段階目が 1,000、2 段階目が「全体のノード数を 500 倍した数値」に設定されており、全体のノード数が 40 の場合は 200,000 回である。また、各ノードがもつベクトルをワード法で分類してクラスター化する、クラスター数は任意に決定できるため、クラスター分析などの結果からクラスター数を調整する。

本研究ではこの KH Coder はデータの前処理段階に使用し、実際の SOM 作成には、データ解析・グラフィックス環境を備えたオープンソースのソフトウェアである R を使用する。KH Coder で出力できる SOM (図 3.11 参照) は抽出語どうしの関係を示すものとなっているため、今回のライフログデータの解析に用いることには向かないためである。

KH Coder で出力できる SOM の R ファイルを基にソースコードを書き換え、R で出力を行う。出力した SOM が図 3.12 になる。この時 SOM 上の数字は id であり、文書同士の関係が表されている。学習回数は 1 段階目が 1,000、2 段階目が 200,000 となっている。また、クラスター数は 6 とした。

図 3.12 より、文書どうしにまとまりは少ないことがわかる。これは文書数が多いことと、対象としているデータが文学作品であることから似たような文章が並ぶことが少ないため

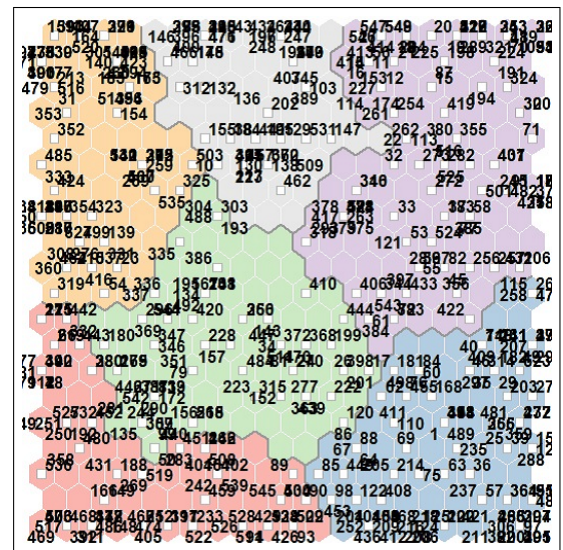
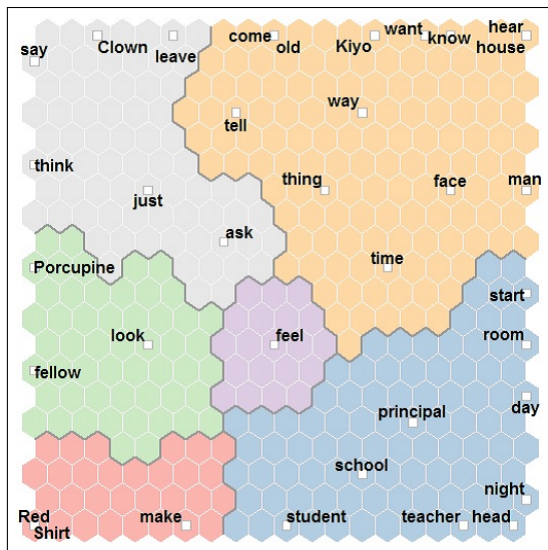


図 3.11: 「坊ちゃん」データから KH Coder 図 3.12: 「坊ちゃん」データから R で作成した SOM
で作成した SOM

データA	データB	データC
- 勉強する - 食事する - 散歩する	- 勉強する - 食事する - 読書する	- 運動する - 運転する - 掃除する

図 3.13: データ A, データ B, データ C の例

であると考えられる。

KH Coder と R を用いてライフログデータであるテキストデータの多変量解析を行い、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やクラスターを構成する抽出語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られる。また、ユーザー自身の複数のライフログデータの中で類似性のあるライフログデータが多ければ、そのライフログデータは平常日を表していると考えられる。一方でライフログデータの類似性ではなく、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、平常日とは違うという危険を察知したりできる。

階層的クラスター分析、共起ネットワークはクラスターの分かれ方やデータを構成する単語の関係性からどのような行動があるのかがわかり、このとき予想できるクラスター数は SOM にも利用できる。MDS、対応分析はプロット点やプロット間隔から類似する行動

やイベント性のある行動がわかる。

SOMはクラスターの分かれ方や時系列を表すプロット順を追っていくことで行動パターンを識別し、多くのライフログの中でも類似したライフログか、特徴的でイベント性のあるライフログかわかる。

図3.13はライフログデータとして、データA, データB, データCがあり、行動識別により各データに三つの行動が存在していることが検出できた場合を表している。この時、データAとデータBは類似性があるといえる。一方でデータA, データBと、データCは類似性がないといえる。この三つのデータが一人のユーザーのライフログデータであれば、平常日とイベント日の比較に利用できる。もし、三つのデータがバラバラのユーザーである場合、ライフログの類似性があるユーザーどうしでコミュニケーションを促進することができたり。ライフログの類似性がないユーザーどうしの比較を行うことで行動の中で改善すべき行動を検出できたりする [38]。このようにライフログデータの類似性やイベント性の検出は様々な応用が可能であると考えられる。

提案手法

§ 4.1 開発システムの概要

本研究で開発するシステムは，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．図 4.1 はシステムの全体図である．まず，データ取得部について提案をする．

本研究では個人情報保護に着目したライフログのため，MOVERIO と画像認識 API を用いたリアルタイム視界情報テキスト変換アプリケーションの開発を行う．開発エンジンは，Unity Technologies が提供するゲーム制作向け開発エンジン Unity5 を使用する．Unity5 は 3D オブジェクトを主として扱い，モバイル端末への出力にも対応している．画像認識 API は Computer Vision API を使用し開発を行う．MOVERIO は Android5.1 であるため API level22 で Android アプリケーションを作成する．

図 4.2 はライフログデータ取得アプリケーションのフローチャートである．起動した際画面は真っ暗であり，カメラを起動しても画面に何も表示を行わないようにしている．MOVERIO は黒い画面は透過する性質があるため，視界を妨げずにライフログデータ取得を行える．本研究では，データが取得できているか常時確認を行うため，取得したタグを邪魔にならない程度の大きさで表示を行うプログラム（ソースコード A.1 参照）を使用している．

カメラ画像を取得すると，画像認識 API へ送信する．画像認識 API を通じて，カメラ画像の情報が JSON データで取得できる．この JSON データに取得時の年月日と時刻を追加して，テキストデータとして MOVERIO 内に保存される．テキストファイルは「long_report_yyyymmdd.txt」という名前で保存され，MOVERIO 内に同じ名前のテキストファイルがなければ新しくテキストファイルを作成し，テキストデータを保存する．同じ名前のテキストファイルがある場合，そのテキストファイルの最後の行にテキストデータを追加し保存する．

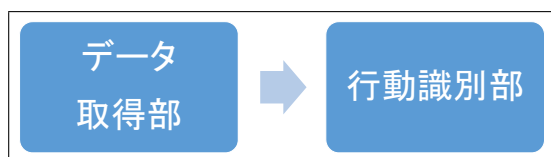


図 4.1: システム全体図

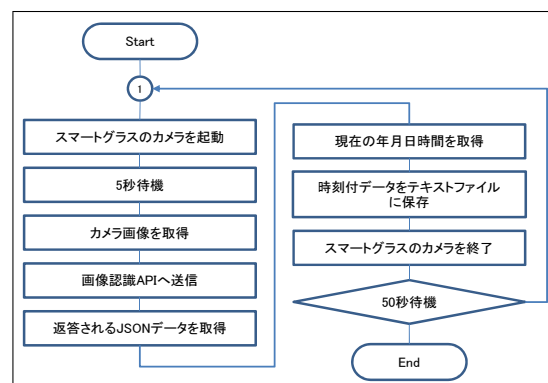


図 4.2: アプリケーションのフローチャート

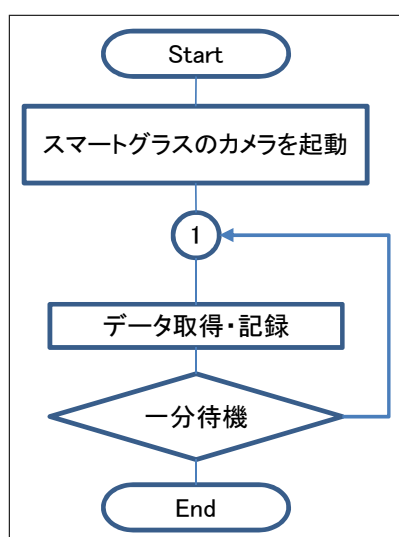


図 4.3: 省電力化前

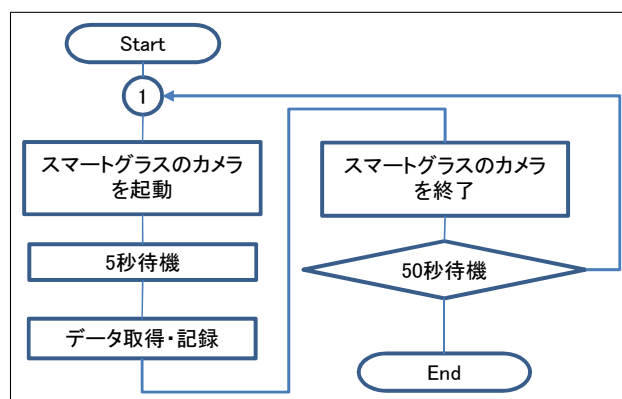


図 4.4: 省電力化後

§ 4.2 省電力化

開発したアプリケーションのバッテリー消耗に関しての工夫について述べる．図 4.3 は、スマートグラスのカメラ機能を起動させたまま 1 分ごとにデータ取得を続けるアプリケーションのフローチャートである．この時、カメラを起動させたままだと 2 時間程度でスマートグラスのバッテリーがなくなってしまう．なお、MOVERIO の標準的な駆動時間は約 6 時間¹³となっている．充電不可能な外出先でのデータ取得のため、少しでも稼働時間を伸ばす必要がある．

この問題に対し、本研究ではカメラの起動・終了にかかるバッテリー消耗よりも、連続起動の方がバッテリー消耗が大きいと考え、データ取得後にカメラ機能を終了するようにプログラムに組み込んでいる（図 4.4 参照）．カメラ終了をプログラムに組み込むことにより、3 時間から 3 時間半程度稼働することができた．

約一分おきにデータを取得するために、カメラの休止時間は 50 秒とし、カメラを立ち上

¹³<http://www.epson.jp/products/moverio/bt300/spec.htm>

げてから5秒後に撮影を行う。理由として、カメラを起動するのに少なからず時間がかかるため、起動後すぐに撮影を行いカメラ画像を取得することは難しいためである。また、その後画像認識 API の応答を得るまでおよそ5秒程度かかるため、約一分ごとにデータを取得できるようにしている。

§ 4.3 周期性の検出

データ取得部の次に行う、行動識別部について述べる。行動識別部では、データ取得部で得たデータを整理し、KH Coder と R を用いて多変量解析を行い、ライフログデータの周期性を検出する。

開発したアプリケーションは、MOVERIO のカメラ画像を画像認識 API に送信し、約一分ごとに以下のテキストデータを取得、記録する。

```
[2018-01-28 10:30:12]{ "description":{ "tags":["indoor","laptop","table","computer","sitting","top","open","desk","white","keyboard","room","man","mouse","plate","laying","bed","playing"], "captions":[{"text":"an open laptop computer sitting on a table","confidence":0.95249546527278339}]},"requestId":"21b3c022-3d1b-4bc1-9cc8-df210d1f2094","metadata":{"height":720,"width":1280,"format":"Jpeg"}}
```

多変量解析を行う前に、前処理としてテキストデータのうち多変量解析に必要なデータのみを抽出する。Computer Vision API はタグとキャプションをライフログデータとして取得できるが、一度に取得するデータが多すぎるとライフログデータとしてノイズとなってしまう。なお、この時の視界は、机の上にノート PC がある状態であり、キャプションの精度は高く、机にあるノート PC を認識できていることがわかる。キャプションだけでは取得できない、indoor などの情報はタグの上位5個に現れていると考え、本研究では tags は confidence の高い順に5個、caption は confidence の最も高いキャプションを使用する。抽出した下記のテキストデータを解析を行いたい時間分テキストファイルに保存する。

```
an open laptop computer sitting on a table,indoor,laptop,table,computer,sitting
```

保存したテキストファイルを KH Coder を使用して、多変量解析する。この時、テキストデータの時系列から周期性を検出するために SOM を使用する。まず、KH Coder で SOM を作成する。このときクラスター数はクラスター解析などの結果から決定できる。KH Coder で SOM を作成した際に取得できる R ファイルの内容を、ライフログデータの時系列関係がわかるように書き換える必要がある。

R ファイルを読み込み、som パッケージを用いて SOM を出力する前に、以下のコマンドを追加する。追加することで、抽出語どうしの関係性を示す SOM ではなく、文章どうし、本研究では一分ごとのライフログデータどうしの関係性を示す SOM を出力できる。

```
d <- t(d)
rownames(d) <- 1:nrow(d)
```

また、本研究ではライフログデータの時系列関係をより視覚的に理解するため、R ファイルの最後に以下のコマンドを追加する。以下のコマンドを追加することで、出力された SOM データが格納されたデータフレーム points がプロットされた id 上に線分のみを上書きできる。

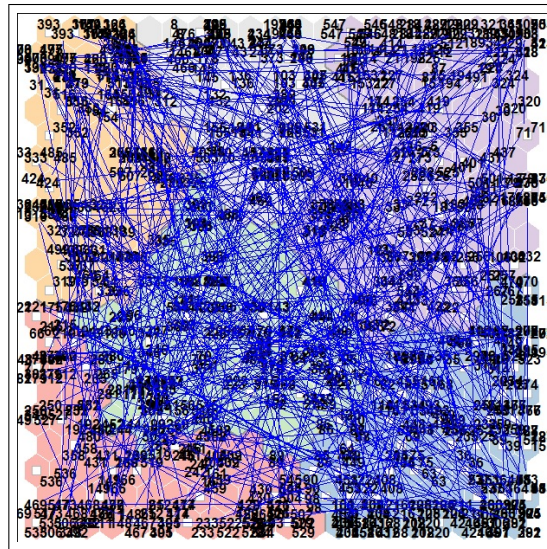


図 4.5: 「坊ちゃん」データから作成した SOM に線分を追加

```
par(new=T)
plot(points[,1],points[,2],type="c",col="色指定")
```

なお、二種類のデータを比較する SOM を作成する場合は、一つのテキストファイルに二種類のテキストファイルをまとめ以下のコマンドを追加する。以下のコマンドを追加することで、データフレーム `points` を二種類のデータに戻し、各々の色で線分を上書きできる。

```
points1<-head(points,n総ライフログデータ=/2)
par(new=T)
plot(points1[,1],points1[,2],type="c",col="色指定 1")

points2<-tail(points,n総ライフログデータ=/2)
par(new=T)
plot(points2[,1],points2[,2],type="c",col="色指定 2")
```

これらの R コマンドを使用して、ライフログデータの時系列を表示できる SOM を出力する。図 4.5 は 3.12 に線分を追加した SOM である。出力された SOM から行動パターンの類似性やイベント性を検出する。

シミュレーション結果ならびに考察

開発したアプリケーションを実際に使用して、ライフログデータを取得する。また、取得したデータを多変量解析を用いて、行動パターンの類似性やイベント性を検出する。

ライフログデータの取得日は2018年1月27日と28日の10時30分から13時30分の180分である。デバイスの充電が100%である状態から充電が切れるまで取得を行ったため取得時間は180分となっている。なお、おおよそ一分に一回データを取得したが、デバイスの処理能力に波があることからデータ数は190となっている。27日に取得したデータをデータ1、28日に取得したデータをデータ2とする。

図5.1にデータ1とデータ2のタイムスケジュールを示す。データ1は学校でデスクトップPCで作業を行い、外出するというライフログデータである。データ2は自宅でノートPCでの作業と食事を行うというライフログデータである。データ1、データ2共に、取得したテキストデータの中から confidence の高いタグ上位5個とキャプションを一行とした190行のテキストファイルを解析に使用する。

データ1とデータ2の比較を行いやすくするため、データ1とデータ2を一つのcsvファイルにしたものをデータ3とする（表5.1参照）。この時label列はデータ1とデータ2の区切りを格納してある。行番号1から190がデータ1であり、191から380がデータ2がとなっている。データ3を用いることで、データ1とデータ2の多変量解析結果を一つの結果上で確認できる。

KH Coder を用いて、取得したテキストファイルの前処理として自然言語処理を行い単語を抽出する。この時、抽出語の中でも、3回以上出現する抽出語を用いる。理由として、データ1、データ2共に抽出語を20個前後にし、プロットされる抽出語を減らすことで解析結果を読み取りやすくするためである。また、解析に使用する品詞は名詞と形容詞に絞る。理由として、動詞は取得したデータ内のキャプションに現れることが多く、コンピューターが置いてある、という状態を an open laptop computer sitting on a table というように画像認識APIが返答してしまうため、本研究ではsittingという状態がノイズになってしまう。そのため品詞を絞り、ノイズを減らすこととした。取得したデータからKH Coderで階層的クラスター分析、MDS、対応分析、共起ネットワーク、SOMを出力する。

まず、データ1とデータ2のクラスター分析を行う。図5.2はデータ1から作成したクラ

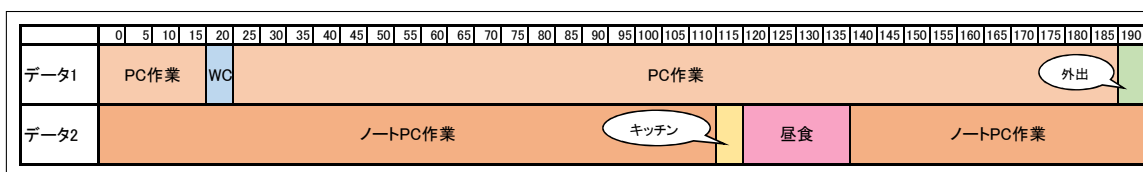


図 5.1: データ 1 とデータ 2 のタイムスケジュール

表 5.1: データ 3

	label	textdata
1	data1	a desk with a computer monitor,indoor,electronics,computer,monitor,table
2	data1	a desk with a computer monitor,indoor,monitor,computer,table,desk
3	data1	a desk with a computer monitor,indoor,computer,table,monitor,desk
4	data1	a desk with a computer monitor,indoor,monitor,table,computer,desk
⋮	⋮	⋮
377	data2	a stack of flyers on a table,indoor,table,top,sitting,desk
378	data2	a stack of flyers on a table,indoor,table,top,sitting,desk
379	data2	a stack of flyers on a table,indoor,table,top,sitting,desk
380	data2	a stack of flyers on a table,indoor,table,top,sitting,desk

スター分析である。この時クラスター数は Auto では 4 となり、図 5.3 の併合標準よりクラスター数 4 前後の傾きに大きな変化がないためクラスター数は 4 のままとした。データ 1 のクラスター分析より、赤のクラスターは computer や keyboard が含まれることから PC 作業であり出現回数もその他のクラスターに比べると最も多いことがわかる。青のクラスターは car や snow が含まれていること、outdoor という単語が含まれていることから外出時のことを表していると考えられる。紫のクラスターは女子トイレの壁の色である white が出現していることからトイレに行くことを示しているように考えた。緑のクラスターは screenshot という単語だけで構成されている、これは視界に画面が大きく含まれている状態ではないかと推測する。

図 5.4 はデータ 2 から作成したクラスター分析である。この時クラスター数は Auto では 5 となり、図 5.5 の併合標準より、クラスター数は 4 よりも 5 が適切であることは明らかのためクラスター数は 5 のままとした。最も多いのは青のクラスターの PC 作業であることが分かる。データ 1 と比較すると、PC 作業を表す単語の中に desktop や keyboard は含まれず、laptop が含まれていることからノート PC での作業を確認できる。ピンクのクラスターは food や plate から食事を表し、緑のクラスターはキッチンを表していると考ええるが、PC で動画を見ながら食事を行っていたため、食べ物以外の物体との関係性が強く表され、自室の私物である flyer や bottle など含まれてしまっている。食事していることをより正確にライフログデータとして取得するには、食べ物をなるべく視界に入れてデータを取得しなくてはいけないのではないかと考えた。また、図 5.2 と図 5.4 を比較すると、データ 2 の自宅の方が視界に入る物体が多いことから bottle や flyer などライフログデータに関係のないものまで取得されていることがわかる。

次に、データ 1 とデータ 2 の MDS を行う。図 5.6 はデータ 1 から作成した MDS である。

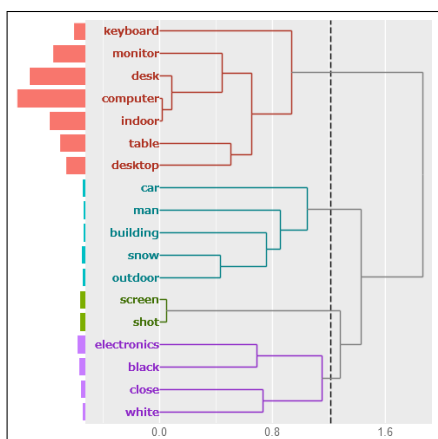


図 5.2: データ 1 のクラスター分析

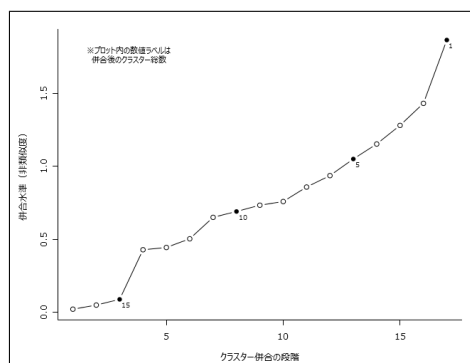


図 5.3: データ 1 の併合標準

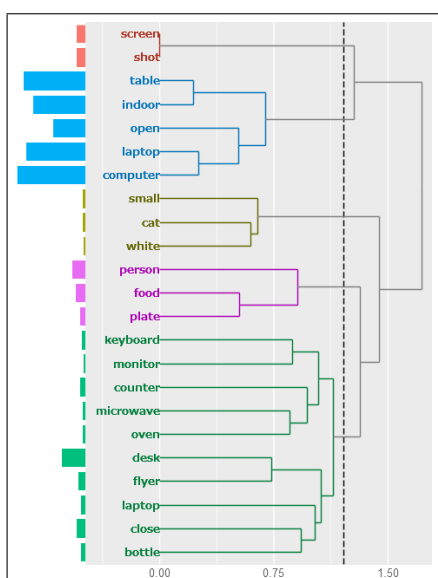


図 5.4: データ 2 のクラスター分析

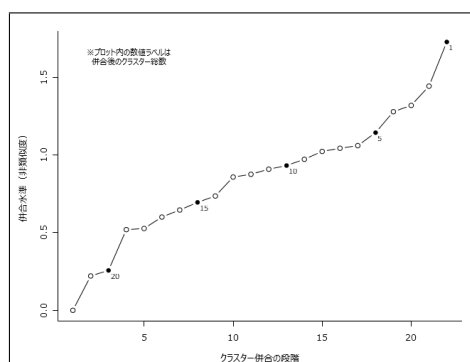


図 5.5: データ 2 の併合標準

クラスター分析より、クラスター数は4とした。PC 作業に関係する computer や desk という抽出語から構成されている一番大きいクラスター 01 と、クラスター 02,03,04 は距離が離れていることとクラスター分けから行動がはっきり分かれていることがわかる。なお、クラスター数を3にして出力を行うとクラスター 03 と 04 が一つのクラスターになったため、クラスター 03 は 02 よりも 04 に近いものとする。

図 5.7 はデータ 2 から作成した MDS である。クラスター分析より、クラスター数は5とした。computer や screen から構成されるクラスター 01 と desk や flyer から構成される 02 は PC 作業という行動を示すため、近い位置に配置されていることがわかる、クラスター 01 と 03 が近いのは、食事の際視界に PC が入り込んでいた影響だと考える。したがって、作業を行いながら食事を行っているのではないかと推測できるプロットとなっている。01 から少し離れた 04 に oven や microwave という抽出語があるため、オーブンや電子レンジを使用したことなどが考えられる。また、クラスター 05 は 01~04 より少し離れているた

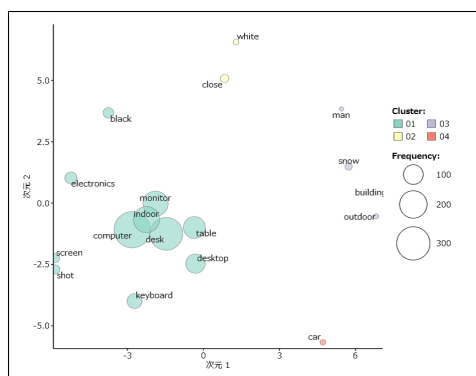


図 5.6: データ 1 の MDS

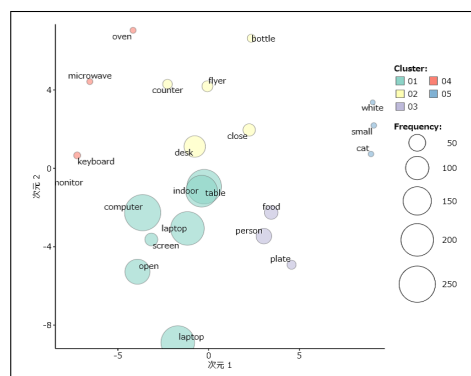


図 5.7: データ 2 の MDS

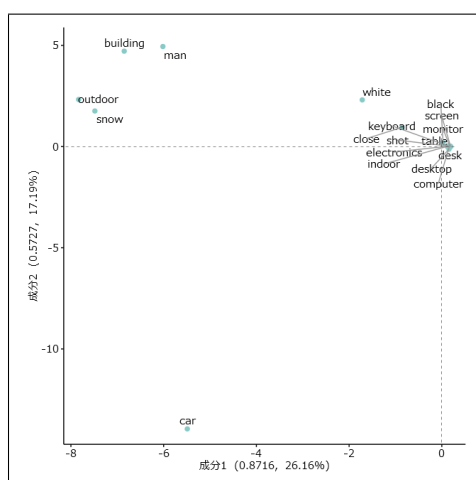


図 5.8: データ 1 の対応分析

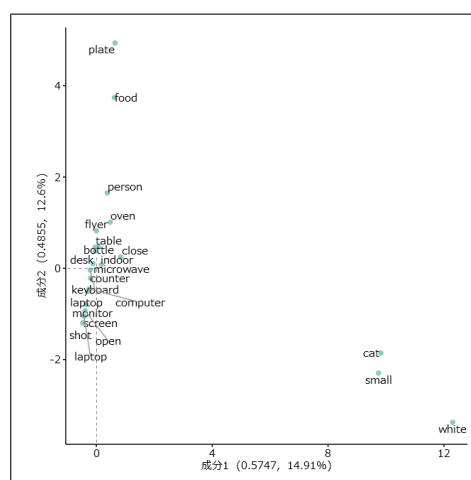


図 5.9: データ 2 の対応分析

めノイズではないかと考えられる．図 5.6 と図 5.7 を比較すると，データ 2 のほうがクラスター間のプロットの距離が近いことがわかる．このことから，実際に似たような行動を複数行っているか，行動を行っている際の視界情報が多いのではないかと考えられる．データ 1 もデータ 2 もクラスター 01 は PC 作業に関する抽出語で構成されているため，PC 作業は類似した行動ではないかと考えられる．クラスター 01 が一番大きく，他のクラスターと大きく差がある点は類似しているが，他のクラスターを構成する抽出語の相違からイベント性を検出できる．

次に，データ 1 とデータ 2 の対応分析を行う．図 5.8 はデータ 1 から作成した対応分析である．図 5.8 より，computer や keyboard で構成される行動である PC 作業を行っていることが最も多く，データ 1 内で特徴的でないことがわかる．また，white や outdoor は原点より離れているため特徴的な行動を構成する抽出語であると考えられる．また，building や outdoor から室外での行動であることが考えられる．

図 5.9 はデータ 2 から作成した対応分析である．図 5.8 と図 5.9 の比較すると，データ 1 もデータ 2 も PC 作業を中心に行っているが，データ 2 は原点より少し離れたところに oven があり，food や white はより特徴的になっていることがわかる．このことから食事をとったことが推測できる．

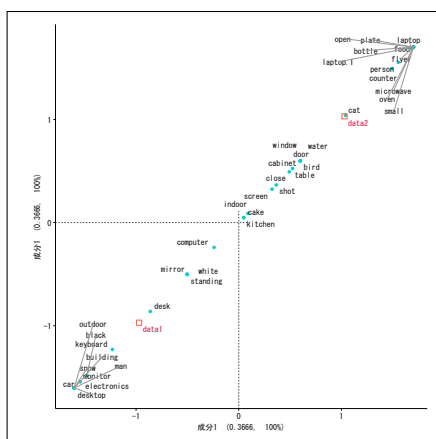


図 5.10: データ 3 の対応分析

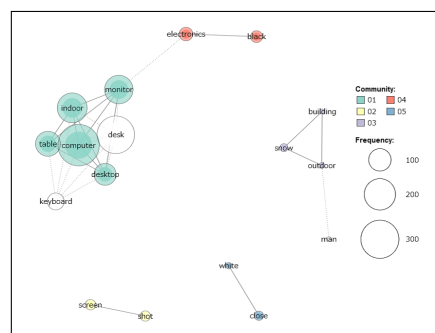


図 5.11: データ 1 の共起ネットワーク

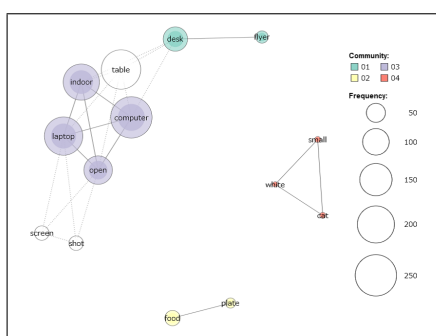


図 5.12: データ 2 の共起ネットワーク

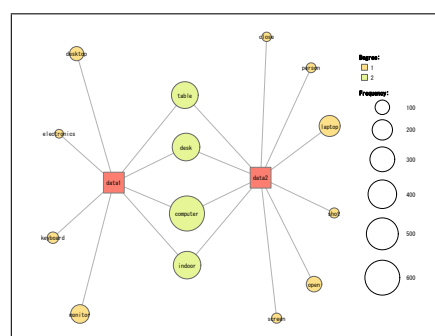


図 5.13: データ 3 の共起ネットワーク

さらに、データ 1 とデータ 2 の関係性を同時に出力することができるため、データ 3 の対応分析を行う。5.10 より、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。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 には類似する行動もあることがわかる。

次に、データ 1 とデータ 2 の共起ネットワークを行う。図 5.11 はデータ 1 から作成した共起ネットワークである。この時、Jaccard 係数が 0.2 以上の共起関係を描画している。図 5.11 より、computer や monitor など PC 作業を表す抽出語どうしは線で結ばれているため、共起関係があることがわかる。また、electronics と black と、クラスターは違っているが共起関係があることがわかる。図 5.12 はデータ 2 から作成した共起ネットワークであり、図 5.11 と比較すると、Jaccard 係数が 0.2 以上の強い共起関係を持つ抽出語が少なく、クラスターも少なくなっていることがわかる。

さらに、データ 1 とデータ 2 の共起関係性を同時に出力するため、データ 3 の共起ネットワークを行う。5.13 より、データ 1 とデータ 2 はともに table, desk, computer, indoor という抽出語と共起関係があり、両方とも Jaccard 係数が 0.2 以上の強い共起関係をもつのは PC 作業を表す抽出語であり、類似性のある行動が確認できる。

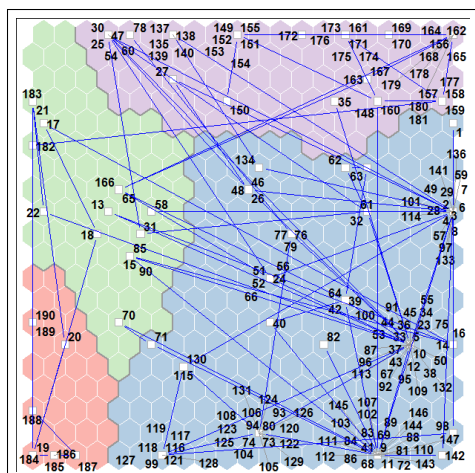


図 5.14: データ 1 の SOM

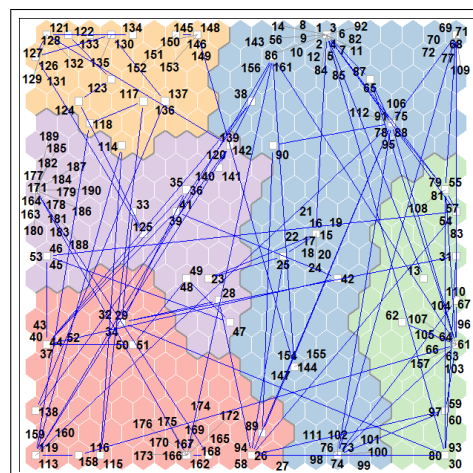


図 5.15: データ 2 の SOM

最後に今までの解析を踏まえてライフログデータの時系列を可視化するため、SOMを作成する。クラスター数はデータ 1 は 4，データ 2 は 5 とした。

図 5.14 はデータ 1 の SOM である。この SOM とテキストデータを照らし合わせると、青のクラスターは PC 作業を表し、緑のクラスターはトイレ、赤のクラスターは外出を表していると考えることができた。なお、外出時もトイレにいる際も視界は白色が多く、white という単語が共通するため、プロットが近いのだと考えた。また、紫のクラスターは a screen shot of a computer や a close up of a computer などの PC 作業の中で出現したノイズのようなテキストから構成されていたが、これ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際に出現するテキストであり、青と紫のクラスターは同じ PC 作業を表している。

図 5.15 はデータ 2 の SOM である。この SOM とテキストデータを同じく照らし合わせると、黄色のクラスターは食事、紫のクラスターは書類が置いてあること、赤のクラスターはキッチンや部屋においてある家具を表していた。青のクラスターは PC 作業をあらわし、緑のクラスターはデータ 1 と同じく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際に出現するテキストから構成されているため、青と緑のクラスターは同じ PC 作業を表している。また、クラスター間を大きくまたぐ線などもあり、常に同じ行動を行っていても、視界に入る物体の変化から線が乱雑になっていると感じた。

データ 3 の SOM を作成し、データ 1 とデータ 2 の時系列の類似性を比較する。データは赤色の線分、データ 2 は青色の線分で示す。クラスター数はデータ 1 とデータ 2 のクラスター数を合わせ 9 とした。また、データ 3 の SOM を作成するソースコードはソースコード A.2 に示す。図 5.16 の、クラスターの分かれ方を 5.17 にしめす。

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られる。また、赤い線と青い線が両方ともつながっている時間のテキストを確認すると、a screen shot of a computer や a close up of a computer などの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や、自宅のポスターを認識していると考えられる。

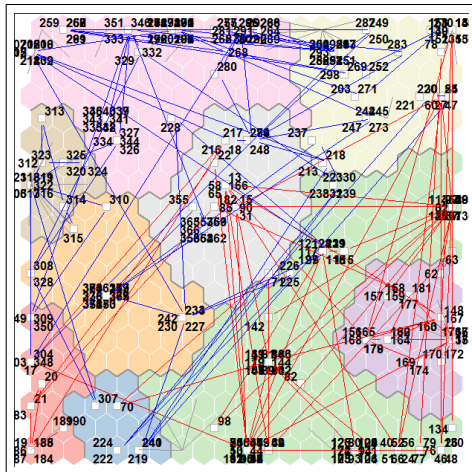


図 5.16: データ 3 の SOM

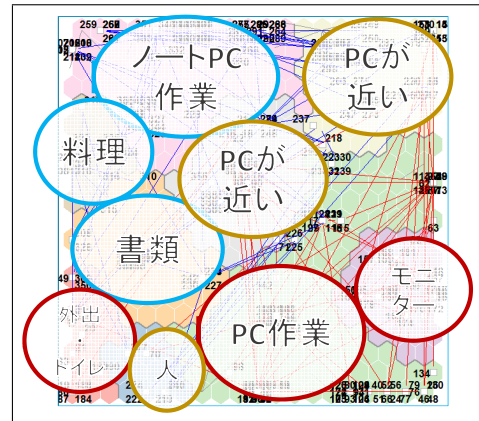


図 5.17: データ 3 の SOM に行動を追記した SOM

階層的クラスター分析，MDS，対応分析，共起ネットワークの解析から，データを構成する行動の検出，類似する行動とそうではないイベント性のある行動を検出することができた．これによって，どのような行動から行っているかという行動識別が可能となっていると考える．また，SOM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るため，GPSを使用せず，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類似性やイベント性を検出できたと考える．

結論と今後の課題

本研究の目的は、多くの人に広く受け入れられるライフログとして、個人情報保護に着目し、手間がかからず自動的にライフログデータの取得を行い、取得したデータから類似性やイベント性を考察できることである。開発したライフログデータ取得アプリケーションを使用したビッグデータ構築・データ解析を行い、行動パターンの類似性・イベント検出を行った。

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、SOMの解析結果より、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。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、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類似性やイベント性を検出できたと考える。本研究の研究成果は、テキストによるライフログデータ取得、解析を行い新たなビジネスプランの検討やユーザー自身の生活の見直しなどに使用できるため、より高度なアプリケーション開発を目指す開発者、研究者の方々の参考になれば幸いである。解明できた点は必ずしも多くはないが、若干なりとも寄与できたと思われる。

今後の課題として、開発したアプリケーションの改善点を上げる。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。組み込むため、取得したいタイミングをMOVERIOに伝える方法の検討も必要となる。

謝辞

本研究を遂行するにあたり，多大なご指導と終始懇切丁寧なご鞭撻を賜った富山県立大学電子・情報工学科の奥原浩之教授に深甚な謝意を表します．最後になりましたが，多大な協力をして頂いた研究室の同輩諸氏に感謝致します．

2018 年 2 月

福嶋 瑞希

参考文献

- [1] 相澤清晴, “ライフログ”, 映像情報メディア学会誌, Vol. 63, No. 4, pp. 445–448, 2009.
- [2] 芳竹宣裕, 伊藤慎, “ユビキタス環境が生み出す大量情報「ライフログ」の活用と実装技術”, NEC 技報, Vol. 62, No. 4, p. 77, 2009.
- [3] 角田宏貴, Hiroki SUMIDA, “ライフログ分析による行動特徴抽出及びイベント検出”, 法政大学大学院紀要 (情報科学研究科編), Vol. 9, pp. 119–124, 2014.
- [4] 矢野裕司, 横井健, 橋山智訓, “行動辞書を利用した Twitter からの行動抽出”, 情報科学技術フォーラム講演論文集, Vol. 11, No. 4, pp. 51–56, 2012.
- [5] 緒方広明, “日本語学習を支援するユビキタス学習環境に関する研究”, <http://www.taf.or.jp/files/items/542/File/P212.pdf>, 閲覧日 2018,1,30.
- [6] 啓之田中, “位置情報の規律のあり方: スマートフォン時代の利便性とプライバシー”, 人間社会研究, Vol. 11, pp. 75–85, 2014.
- [7] 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.
- [8] 北村圭吾, 山崎俊彦, 相澤清晴, “食事ログの取得と処理ー画像処理による食事記録ー”, 映像情報メディア学会誌, Vol. 63, No. 3, pp. 376–379, 2009.
- [9] 株式会社 N T T データ経営研究所, “日本語学習を支援するユビキタス学習環境に関する研究”, <http://www.keieiken.co.jp/aboutus/newsrelease/161122/>, 閲覧日 2018,2,5.
- [10] 入江英嗣, 森田光貴, 岩崎央, 千竈航平, 放地宏佳, 小木真人, 樫原裕大, 芝星帆, 眞島一貴, 努吉永, “AirTarget: 光学シースルー方式 HMD とマーカレス画像認識による高可搬性実世界志向インタフェース”, 情報処理学会論文誌, Vol. 55, No. 4, pp. 1415–1427, 2014.
- [11] 川上晃平, “スマートグラスを利用した授業支援システムの開発”, 2017.
- [12] 倉田陽平, 真田風, 鈴木祥平, 石川博, “Flickr と Google Cloud Vision API によりテーマ別観光マップを作る試み”, <http://db-event.jpn.org/deim2017/papers/321.pdf>, 閲覧日 2018,1,4.
- [13] 大雄治, 吉川眞, 田中一成, “ソーシャルメディアを活用した景観の分析と評価”, 日本都市計画学会関西支部研究発表会講演概要集, Vol. 15, pp. 13–16, 2017.
- [14] 小林亜令, 岩本健嗣, 西山智, “釈迦: 携帯電話を用いたユーザ移動状態推定・共有方式-モバイルコンピューティング, モバイルアプリケーション, ユビキタス通信, モバイルマルチメディア通信-”, 電子情報通信学会技術研究報告. MoMuC, モバイルマルチメディア通信, Vol. 108, No. 44, pp. 115–120, 2008.

- [15] 寺田努, “ウェアラブルセンサを用いた行動認識技術の現状と課題”, コンピュータ ソフトウェア, Vol. 28, No. 2, pp. 43–54, 2011.
- [16] 貴志一樹, 山崎俊彦, 相澤清晴, “机上行動のライフログのための行動認識”, 映像情報メディア学会年次大会講演予稿集, Vol. 2014, , 2014.
- [17] 前川卓也, 柳沢豊, 岸野泰恵, 石黒勝彦, 亀井剛次, 櫻井保志, 岡留剛, “ウェアラブルセンサによるモノを用いた行動の認識について”, Technical Report 57, 研究報告ユビキタスコンピューティングシステム (UBI) , 2010.
- [18] 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.
- [19] 佐野香織, 李在鎬, “KH Coder で何ができるか: 日本語習得・日本語教育研究利用への示唆”, 言語文化と日本語教育, Vol. 33, pp. 94–95, 2007.
- [20] “KH Coder を用いた研究事例のリスト”, <http://khc.sourceforge.net/bib.html>, 閲覧日 2018,1,7.
- [21] 二宮隆次, 小野浩幸, 高橋幸司, 野田博行, “新聞記事を基にしたテキストマイニング手法による産学官連携活動分析”, 科学・技術研究, Vol. 5, No. 1, pp. 93–104, 2016.
- [22] “クラスター分析の手法 (階層クラスター分析) — データ分析基礎知識”, https://www.albert2005.co.jp/knowledge/data_mining/cluster/hierarchical_clustering, 閲覧日 2018,1,24.
- [23] “階層的クラスター分析について”, http://koichi.nihon.to/cgi-bin/bbs_khn/khcf.cgi?list=&no=977&mode=allread&page=0, 閲覧日 2018,1,24.
- [24] 吉原一紘, 徳高平蔵, “クラスター分析の概要”, *Journal of Surface Analysis*, Vol. 21, No. 1, pp. 10–17, 2014.
- [25] 李美龍, 田中恒也, 成田吉弘, “画像を用いた製品の「飽き」に関する感性評価: デザインの視覚的要素を中心に—”, 日本感性工学会論文誌, Vol. 11, No. 3, pp. 407–417, 2012.
- [26] 小峯敦・下平裕之, “ベヴァリッジ『自由社会における完全雇用』のケインズの要素-テキストマイニングを加味した量的・質的分析-”, 2017.
- [27] 齋藤堯幸, “多次元尺度構成法”, 計測と制御, Vol. 22, No. 1, pp. 126–131, 1983.
- [28] Joseph B Kruskal, “Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis”, *Psychometrika*, Vol. 29, No. 1, pp. 1–27, 1964.
- [29] “Jaccard 係数の計算式と特徴 (1) ”, <https://www.slideshare.net/khcoder/jaccard1>, 閲覧日 2018,2,3.

- [30] 中山慶一郎, “< 研究ノート > 対応分析によるデータ解析”, 関西学院大学社会学部紀要, No. 108, pp. 133–145, 2009.
- [31] 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.
- [32] “共起ネットワークにおける中心性の解釈について”, http://www.koichi.nihon.to/cgi-bin/bbs_khn/khcf.cgi?no=2493&mode=allread#2496, 閲覧日 2018,2,2.
- [33] 横田尚己, 山田圭二郎, “熊本地震のつぶやきに見る感情極性値の時空間解析”, 都市計画論文集, Vol. 52, No. 3, pp. 1081–1087, 2017.
- [34] 増田正, Masuda Tadashi, 高崎経済大学地域政策学部, “地方議会の会議録に関するテキストマイニング分析：高崎市議会を事例として”, 地域政策研究 = Studies of regional policy, Vol. 15, No. 1, pp. 17–31, 2012.
- [35] T. KOHONEN, “Self-organized formation of topologically correct feature map”, *Biol. Cybern.*, Vol. 43, pp. 59–69, 1982.
- [36] 岡晋之介, “自己組織化マップを用いた気象要素の分類と予測”, <http://www.gifu-nct.ac.jp/elec/deguchi/sotsuron/oka/oka.html>, 閲覧日 2018,1,7.
- [37] “自己組織化特徴マップ (SOM) ”, <http://www.sist.ac.jp/kanakubo/research/neuro/selforganizingmap.html>, 閲覧日 2018,1,31.
- [38] 勝治宏基, 米澤拓郎, 中澤仁, 高汐一紀, 徳田英幸ほか, “Synchrometer: ライフログを利用した日常行動における他者との類似度生成”, 研究報告ユビキタスコンピューティングシステム (UBI), Vol. 2013, No. 17, pp. 1–7, 2013.

付録

A. 1 ライフログデータ取得アプリケーションのソースコード

ライフログデータ取得アプリケーションのソースコード A.1 をしめす.

ソースコード A. 1: app.cs

```
1 using UnityEngine;
2 using System.Collections;
3 using System.Collections.Generic;
4 using UnityEngine.UI;
5 using System.IO;
6 using System;
7 using System.Text;
8 using System.Linq;
9
10 public class som : MonoBehaviour
11 {
12     private float captureIntervalSeconds = 50.0f;
13     private float captureIntervalSeconds2 = 5.0f;
14     public Text gtext;
15     Dictionary<string, string> headers;
16     private int Width = 1280;
17     private int Height = 720;
18     private int FPS = 30;
19     private WebCamTexture webcamTexture;
20     private Color32[] color32;
21     private string tagsStg;
22     private string reportFileName2 = "long_report.txt";
23     public bool addDateTime = true;
24
25     void Start ()
26     {
27         Screen.sleepTimeout = SleepTimeout.NeverSleep;
28         StartCoroutine ("Sample");
29     }
30
31     public IEnumerator Sample ()
32     {
33         WebCamDevice[] devices = WebCamTexture.devices;
34         WebCamDevice userCameraDevice = WebCamTexture.devices [0];
35         webcamTexture = new WebCamTexture (userCameraDevice.name,
36             Width, Height, FPS);
37         webcamTexture.Play ();
38         Debug.Log ("webcamTexture");
39
40         yield return new WaitForSeconds (captureIntervalSeconds2);
41         color32 = webcamTexture.GetPixels32 ();
42         Texture2D texture = new Texture2D (webcamTexture.width,
43             webcamTexture.height);
44         texture.SetPixels32 (color32);
45         texture.Apply ();
46         byte[] jpg = texture.EncodeToJPG ();
```

```

45     string VISIONKEY = "ba7982e18b4943d18024749aca8031fb";
46     var uri = "https://westus.api.cognitive.microsoft.com/vision/
        v1.0/describe";
47
48     string responseData;
49     var headers = new Dictionary<string, string> () {
50         { "Ocp-Apim-Subscription-Key", VISIONKEY },
51         { "Content-Type", "application/octet-stream" }
52     };
53
54     WWW www = new WWW (uri, jpg, headers);
55     yield return www;
56     responseData = www.text;
57     var data = JsonUtility.FromJson<SumDataFromJson> (responseData);
58
59     DateTime dt = DateTime.Now;
60     string text2 = dt.ToString (" [yyyy-MM-dd_HH:mm:ss] ") +
        responseData.ToString () + "\n";
61     string outfile2 = reportFileName2;
62
63     if (addDateTime) {
64         string file2 = Path.GetFileNameWithoutExtension (
            reportFileName2);
65         string ext2 = Path.GetExtension (reportFileName2);
66         outfile2 = file2 + "_" + dt.ToString ("yyyyMMdd") +
            ext2;
67     }
68
69     SaveText (text2, Path.Combine (Application.persistentDataPath,
        outfile2));
70     MonoBehaviour.Destroy (texture);
71     MonoBehaviour.Destroy (webcamTexture);
72     color32 = null;
73     StartCoroutine ("StopRunTimeTemp");
74 }
75
76 public IEnumerator StopRunTimeTemp ()
77 {
78     webcamTexture.Stop ();
79     gtext.text = tagsStg;
80     yield return new WaitForSeconds (captureIntervalSeconds);
81     StartCoroutine ("Sample");
82 }
83
84 public static bool SaveText (string text, string path)
85 {
86     try {
87         using (StreamWriter writer = new StreamWriter (path, true))
88         {
89             writer.Write (text);
90             writer.Flush ();
91             writer.Close ();
92         }
93     } catch (Exception e) {
94         Debug.Log (e.Message);
95         return false;
96     }
97     return true;
98 }

```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

408 rownames(d) <- 1:nrow(d)
409 # SOM
410 library(som)
411 somm <- som(
412     d,
413     n_nodes,
414     n_nodes,
415     topol="hexa",
416     rlen=c(rlen1,rlen2)
417 )
418
419 word_labs <- rownames(d)
420 n_words <- length(word_labs)
421
422
423 color_universal_design <- 1
424 # END: DATA
425
426 cex <- 1
427 text_font <- 2
428 if_cls <- 1
429 n_cls <- 9
430 if_plothex <- 1
431
432 # n_nodes <- 20
433 # rlen1 <- 1000
434 # rlen2 <- 200000
435
436
437
438
439
440 row2coods <- NULL
441 eve <- 0
442 for (i in 0:(n_nodes - 1)){
443     for (h in 0:(n_nodes - 1)){
444         row2coods <- c(row2coods, h + eve, i)
445     }
446     if (eve == 0){
447         eve <- 0.5
448     } else {
449         eve <- 0
450     }
451 }
452 row2coods <- matrix( row2coods, byrow=T, ncol=2 )
453
454
455 if ( if_cls == 1 ){
456     library( RColorBrewer )
457
458     if (
459         ( as.numeric( R.Version()$major ) >= 3 )
460         && ( as.numeric( R.Version()$minor ) >= 1.0)
461     ){ # >= R 3.1.0
462         hcl <- hclust( dist(somm$code,method="euclidean"), method="ward
463             .D2" )
464     } else { # <= R 3.0
465         hcl <- hclust( dist(somm$code,method="euclidean")^2, method="
466             ward" )
467     }

```

```

466
467     colors <- NULL
468     if (n_cls <= 9){
469         pastel <- brewer.pal(9, "Pastel1")
470         # pastel[6] = brewer.pal(9, "Pastel1")[9]
471         # pastel[9] = brewer.pal(9, "Pastel1")[6]
472         pastel[6] = "gray91"
473         pastel[9] = "#F5F5DC" # FAF3C8 F7F1C6 EEE8AA F0E68C
474         colors <- pastel[cutree(hcl,k=n_cls)]
475     }
476     if (n_cls > 9) {
477
478         library(colorspace)
479         new_col <- order( runif(n_cls) )
480         colors <-
481             rainbow_hcl(n_cls, start=20, end=340, l=92, c=20)[
482                 #terrain_hcl(n_cls, c = c(35, 5), l = c(85, 95), power = c(0.5,1)
483                 )/
484                 new_col[cutree(hcl,k=n_cls)]
485     ]
486 } else {
487     colors <- rep("gray90", n_nodes^2)
488 }
489 labcd <- NULL
490
491
492
493
494 plot_mode <- "color"
495
496
497 par(mai=c(0,0,0,0), mar=c(0,0,0,0), omi=c(0,0,0,0), oma =c(0,0,0,0) )
498
499 plot(
500     NULL,NULL,
501     xlim=c(0,n_nodes-0.5),
502     ylim=c(0,n_nodes-1),
503     axes=F,
504     frame.plot=F
505 )
506
507 if (if_plothe == 1){
508     a <- 0.333333333333
509 } else {
510     a <- 0.5
511 }
512 b <- 1-a
513
514 color_pte <- "gray70"
515 cls_lwd <- 2
516
517 if ( plot_mode == "gray"){
518     color_act <- rep("white",n_nodes^2)
519     if_points <- 1
520     w_lwd <- 1
521     cls_lwd <- 2.25
522     color_cls <- "gray35"
523     color_line <- "gray50"
524     color_pte <- "gray40"

```

```

525     color_ptf <- "gray85"
526   }
527   if ( plot_mode == "color" ) {
528     color_act <- colors
529     color_line <- "white"
530     if_points <- 1
531     w_lwd <- 1
532     if (n_cls > 9) {
533       color_cls <- "gray45"
534     } else {
535       color_cls <- "gray60"
536     }
537     color_ptf <- "white"
538   }
539   if ( plot_mode == "freq" ){
540     color_act <- somm$code.sum$noobs;
541     if (max(color_act) == 1){
542       color_act <- color_act * 3 + 1;
543     } else {
544       color_act <- color_act - min(color_act)
545       color_act <- round( color_act / max(color_act) * 6 ) + 1
546       #color_act[color_act==7] <- 6
547     }
548     color_seed <- brewer.pal(6,"GnBu")
549     #color_seed <- brewer.pal(6,"YlOrRd")
550     color_seed <- c("white", color_seed)
551     color_act <- color_seed[color_act]
552
553     color_line <- "gray70"
554     if_points <- 0
555     w_lwd <- 1
556     color_cls <- "gray45"
557     color_ptf <- "white"
558   }
559   if ( plot_mode == "umat" ){
560
561
562     dist_u <- NULL
563
564     dist_m <- as.matrix( dist(somm$code, method="euclid") )
565
566     for (i in 0:(n_nodes - 1)){
567       for (h in 0:(n_nodes - 1)){
568         cu <- NULL
569         n <- 0
570
571         if (h != n_nodes - 1){
572           cu <- c(
573             cu,
574             dist_m[
575               h + i * n_nodes + 1,
576               h + 1 + i * n_nodes + 1
577             ]
578           )
579         }
580
581         if (h != 0){
582           cu <- c(
583             cu,
584             dist_m[

```

```

585             h + i * n_nodes + 1,
586             h - 1 + i * n_nodes + 1
587         ]
588     )
589 }
590
591 if (i != n_nodes - 1){
592     if (h %% 2 == 0){
593         cu <- c(
594             cu,
595             dist_m[
596                 h + i * n_nodes + 1,
597                 h + (i + 1) * n_nodes + 1
598             ]
599         )
600     } else {
601         if (h != n_nodes - 1){
602             cu <- c(
603                 cu,
604                 dist_m[
605                     h + i * n_nodes + 1,
606                     h + 1 + (i + 1) * n_
607                         nodes + 1
608                 ]
609             )
610         }
611     }
612 }
613
614 if (i != 0){
615     if (h %% 2 == 0){
616         cu <- c(
617             cu,
618             dist_m[
619                 h + i * n_nodes + 1,
620                 h + (i - 1) * n_nodes + 1
621             ]
622         )
623     } else {
624         if (h != n_nodes - 1){
625             cu <- c(
626                 cu,
627                 dist_m[
628                     h + i * n_nodes + 1,
629                     h + 1 + (i - 1) * n_
630                         nodes + 1
631                 ]
632             )
633         }
634     }
635 }
636
637 if (i != n_nodes - 1){
638     if (h %% 2 == 0){
639         if (h != 0){
640             cu <- c(
641                 cu,
642                 dist_m[
643                     h + i * n_nodes + 1,

```

```

642                                     h - 1 + ( i + 1 ) * n_
                                                nodes + 1
643                                 ]
644                             )
645                         }
646                     } else {
647                         cu <- c(
648                             cu,
649                             dist_m[
650                                 h + i * n_nodes + 1,
651                                 h + ( i + 1 ) * n_nodes + 1
652                             ]
653                         )
654                     }
655                 }
656             if (i != 0){
657                 if (h %% 2 == 0){
658                     if (h != 0){
659                         cu <- c(
660                             cu,
661                             dist_m[
662                                 h + i * n_nodes + 1,
663                                 h - 1 + ( i - 1 ) * n_
664                                     nodes + 1
665                             ]
666                         )
667                     }
668                 } else {
669                     cu <- c(
670                         cu,
671                         dist_m[
672                             h + i * n_nodes + 1,
673                             h + ( i - 1 ) * n_nodes + 1
674                         ]
675                     )
676                 }
677             }
678             dist_u <- c(dist_u, median(cu) )
679         }
680     }
681
682     print( summary(dist_u) )
683
684     dist_u <- dist_u - min(dist_u)
685     dist_u <- round( dist_u / max(dist_u) * 100 ) + 1
686
687     if (color_universal_design == 0){
688         color_act <- cm.colors(101)[dist_u]
689         color_line <- "gray70"
690         color_cls <- "gray45"
691     } else {
692         library(RColorBrewer)
693         if (T){
694             col_seed <- brewer.pal(9, "GnBu")
695             myPalette <- colorRampPalette( col_seed )
696             color_act <- myPalette(dist_u)
697             color_act <- adjustcolor(color_act, alpha=0.8)
698             color_line <- "white"

```

```

699         color_cls <- "gray30"
700     } else {
701         col_seed <- rev(brewer.pal(9, "RdYlBu"))
702         myPalette <- colorRampPalette( col_seed )
703         color_act <- myPalette(101)[dist_u]
704         color_act <- adjustcolor(color_act, alpha=0.8)
705         color_line <- "gray50"
706         color_cls <- "gray25"
707     }
708 }
709
710 #color_line <- "gray70"
711 if_points <- 1
712 w_lwd <- 1
713 #color_cls <- "gray45"
714 color_ptf <- "white"
715 }
716
717
718 for (i in 1:n_nodes^2){
719     x <- row2coods[i,1]
720     y <- row2coods[i,2]
721
722     polygon(
723         x=c( x + 0.5, x + 0.5, x, x - 0.5, x - 0.5, x ),
724         y=c( y + a, y - a, y - b, y - a, y + a, y + b ),
725         col=color_act[i],
726         border="white",
727         lty=0,
728     )
729 }
730
731
732 for (i in 0:(n_nodes - 1)){
733     for (h in 0:(n_nodes - 2)){
734         if ( colors[h + i * n_nodes + 1] == colors[h + i * n_nodes + 2] ){
735             x <- h
736             y <- i
737             if ( y %% 2 == 1 ){
738                 x <- x + 0.5
739             }
740
741             segments(
742                 x + 0.5, y + a,
743                 x + 0.5, y - a,
744                 col=color_line,
745                 lwd=w_lwd,
746             )
747         }
748     }
749 }
750
751 for (i in 0:(n_nodes - 1)){
752     for (h in c(-1, n_nodes-1) ){
753         x <- h
754         y <- i
755         if ( y %% 2 == 1 ){
756             x <- x + 0.5
757         }
758         segments(

```

```

759             x + 0.5, y + a,
760             x + 0.5, y - a,
761             col=color_line,
762             lwd=w_lwd,
763         )
764     }
765     if ( y %% 2 == 0 ){
766         segments(
767             -0.5, y + a,
768             0 , y + 1 - a,
769             col=color_line,
770             lwd=w_lwd,
771         )
772         if ( y != 0){
773             segments(
774                 -0.5, y - a,
775                 0 , y - 1 + a,
776                 col=color_line,
777                 lwd=w_lwd,
778             )
779         }
780     } else {
781         if ( y != n_nodes - 1){
782             segments(
783                 n_nodes - 0.5, y + 1 - a,
784                 n_nodes , y + a,
785                 col=color_line,
786                 lwd=w_lwd,
787             )
788         }
789         segments(
790             n_nodes - 0.5, y - 1 + a,
791             n_nodes , y - a,
792             col=color_line,
793             lwd=w_lwd,
794         )
795     }
796 }
797
798 for (i in 0:(n_nodes - 2)){
799     for (h in 0:(n_nodes - 1)){
800         if (i %% 2 == 1){
801             chk <- 1
802         } else {
803             chk <- 0
804         }
805
806         if (
807             is.na(colors[h + i * n_nodes + 1]) == 1
808             || is.na(colors[h + chk + (i+1) * n_nodes + 1]) == 1
809             || h + chk == n_nodes
810         ){
811             next
812         }
813
814         if (
815             colors[h + i * n_nodes + 1]
816             == colors[h + chk + (i+1) * n_nodes + 1]
817         ){
818             x <- h

```

```

819         y <- i
820         if ( y %% 2 == 1 ){
821             x <- x + 0.5
822         }
823
824         segments(
825             x, y + b,
826             x + 0.5, y + a,
827             col=color_line,
828             lwd=w_lwd,
829         )
830     }
831 }
832 }
833
834 for (i in 0:(n_nodes - 2)){
835     for (h in 0:(n_nodes - 1)){
836         if (i %% 2 == 0){
837             chk <- 1
838         } else {
839             chk <- 0
840         }
841         if (
842             is.na(colors[h + i * n_nodes + 1]) == 1
843             || is.na(colors[h - chk + (i+1) * n_nodes + 1]) == 1
844             || h - chk < 0
845         ){
846             next
847         }
848
849         if (
850             colors[h + i * n_nodes + 1]
851             == colors[h - chk + (i+1) * n_nodes + 1]
852         ){
853             x <- h
854             y <- i
855             if ( y %% 2 == 1 ){
856                 x <- x + 0.5
857             }
858
859             segments(
860                 x, y + b,
861                 x - 0.5, y + a,
862                 col=color_line,
863                 lwd=w_lwd,
864             )
865         }
866     }
867 }
868
869
870
871 for (i in 0:(n_nodes - 1)){
872     for (h in 0:(n_nodes - 2)){
873         if ( colors[h + i * n_nodes + 1] != colors[h + i * n_nodes + 2] ){
874             x <- h
875             y <- i
876             if ( y %% 2 == 1 ){
877                 x <- x + 0.5

```

```

878         }
879
880         segments(
881             x + 0.5, y + a,
882             x + 0.5, y - a,
883             col=color_cls,
884             lwd=cls_lwd,
885         )
886     }
887 }
888 }
889
890 for (i in 0:(n_nodes - 2)){
891     for (h in 0:(n_nodes - 1)){
892         if (i %% 2 == 1){
893             chk <- 1
894         } else {
895             chk <- 0
896         }
897
898         if (
899             is.na(colors[h + i * n_nodes + 1]) == 1
900             || is.na(colors[h + chk + (i+1) * n_nodes + 1]) == 1
901             || h + chk == n_nodes
902         ){
903             next
904         }
905
906         if (
907             colors[h + i * n_nodes + 1]
908             != colors[h + chk + (i+1) * n_nodes + 1]
909         ){
910             x <- h
911             y <- i
912             if ( y %% 2 == 1 ){
913                 x <- x + 0.5
914             }
915
916             segments(
917                 x, y + b,
918                 x + 0.5, y + a,
919                 col=color_cls,
920                 lwd=cls_lwd,
921             )
922         }
923     }
924 }
925
926 for (i in 0:(n_nodes - 2)){
927     for (h in 0:(n_nodes - 1)){
928         if (i %% 2 == 0){
929             chk <- 1
930         } else {
931             chk <- 0
932         }
933
934         if (
935             is.na(colors[h + i * n_nodes + 1]) == 1
936             || is.na(colors[h - chk + (i+1) * n_nodes + 1]) == 1

```

```

937         || h - chk < 0
938     ){
939         next
940     }
941
942     if (
943         colors[h + i * n_nodes + 1]
944         != colors[h - chk + (i+1) * n_nodes + 1]
945     ){
946         x <- h
947         y <- i
948         if ( y %% 2 == 1 ){
949             x <- x + 0.5
950         }
951
952         segments(
953             x, y + b,
954             x - 0.5, y + a,
955             col=color_cls,
956             lwd=cls_lwd,
957         )
958     }
959 }
960 }
961
962 points <- NULL
963 sf <- 0.35
964 a <- a * sf;
965 b <- b * sf;
966 c <- 0.5 * sf;
967 for (i in 1:nrow(somm$visual)){
968     x <- somm$visual[i,1]
969     y <- somm$visual[i,2]
970     if ( y %% 2 == 1 ){
971         x <- x + 0.5
972     }
973     points <- c(points, x, y)
974 }
975 points <- matrix( points, byrow=T, ncol=2 )
976
977 if( if_points == 1 ){
978     if (F){
979         for (i in 1:nrow(points)){
980             x <- points[i,1]
981             y <- points[i,2]
982
983             polygon(
984                 x=c( x + c, x + c, x, x - c, x - c, x ),
985                 y=c( y + a, y - a, y - b, y - a, y + a, y + b ),
986                 col=color_ptf,
987                 border=color_pte,
988                 lty=1,
989             )
990         }
991     } else {
992         symbols(
993             points[,1],
994             points[,2],
995             squares=rep(0.35,length(points[,1])),

```

```

996             #circles=rep(0.2,length(points[,1])),
997             fg="gray70",
998             bg=color_ptf,
999             inches=F,
1000             add=T,
1001         )
1002     }
1003 }
1004
1005 library(maptools)
1006 if (is.null(labcd) == 1){
1007     labcd <- pointLabel(
1008         x=points[,1],
1009         y=points[,2],
1010         labels=word_labs,
1011         doPlot=F,
1012         cex=cex,
1013         offset=0
1014     )
1015
1016     xorg <- points[,1]
1017     yorg <- points[,2]
1018     #cex <- 1
1019
1020     if ( length(xorg) < 300 ) {
1021         library(wordcloud)
1022
1023         # fix for "wordlayout" function
1024         filename <- tempfile()
1025         writeLines("wordlayout <- function (x, y, words, cex = 1, rotate90 = FALSE,
1026             xlim = c(-Inf,
1027                 Inf), ylim = c(-Inf, Inf), tstep = 0.1, rstep = 0.1, ...)
1028         {
1029             tails <- "\"g|j|p|q|y\"
1030             n <- length(words)
1031             sdx <- sd(x, na.rm = TRUE)
1032             sdy <- sd(y, na.rm = TRUE)
1033             iterations <- 0
1034             if (sdx == 0)
1035                 sdx <- 1
1036             if (sdy == 0)
1037                 sdy <- 1
1038             if (length(cex) == 1)
1039                 cex <- rep(cex, n)
1040             if (length(rotate90) == 1)
1041                 rotate90 <- rep(rotate90, n)
1042             boxes <- list()
1043             for (i in 1:length(words)) {
1044                 rotWord <- rotate90[i]
1045                 r <- 0
1046                 theta <- runif(1, 0, 2 * pi)
1047                 x1 <- xo <- x[i]
1048                 y1 <- yo <- y[i]
1049                 wid <- strwidth(words[i], cex = cex[i], ...)
1050                 ht <- strheight(words[i], cex = cex[i], ...)
1051                 if (grepl(tails, words[i]))
1052                     ht <- ht + ht * 0.2
1053                 if (rotWord) {
1054                     tmp <- ht

```

```

1055         ht <- wid
1056         wid <- tmp
1057     }
1058     isOverlaped <- TRUE
1059     while (isOverlaped) {
1060         if (!.overlap(x1 - 0.5 * wid, y1 - 0.5 * ht, wid,
1061             ht, boxes) && x1 - 0.5 * wid > xlim[1] && y1
1062             - 0.5 * ht > ylim[1] && x1 + 0.5 * wid < xlim[2]
1063             && y1 + 0.5 * ht < ylim[2]) {
1064             boxes[[length(boxes) + 1]] <- c(x1 - 0.5 * wid
1065                 , y1 - 0.5 * ht, wid, ht)
1066             isOverlaped <- FALSE
1067         }
1068         else {
1069             theta <- theta + tstep
1070             r <- r + rstep * tstep / (2 * pi)
1071             x1 <- xo + sdx * r * cos(theta)
1072             y1 <- yo + sdy * r * sin(theta)
1073             iterations <- iterations + 1
1074             if (iterations > 500000){
1075                 boxes[[length(boxes) + 1]] <- c(x1 -
1076                     0.5 * wid,
1077                     y1 - 0.5 * ht, wid, ht)
1078                 isOverlaped = FALSE
1079             }
1080         }
1081     }
1082     print( paste("\ iterations: \", iterations) )
1083     result <- do.call(rbind, boxes)
1084     colnames(result) <- c("x", "y", "width", "ht")
1085     rownames(result) <- words
1086     result
1087 }
1088 ", filename)
1089 insertSource(filename, package="wordcloud", force=FALSE)
1090 nc <- wordlayout(
1091     labcd$x,
1092     labcd$y,
1093     word_labs,
1094     cex=cex * 1.25,
1095     xlim=c( par( "usr" )[1], par( "usr" )[2] ),
1096     ylim=c( par( "usr" )[3], par( "usr" )[4] )
1097 )
1098
1099 xlen <- par("usr")[2] - par("usr")[1]
1100 ylen <- par("usr")[4] - par("usr")[3]
1101
1102 segs <- NULL
1103 for (i in 1:length(word_labs) ){
1104     x <- ( nc[i,1] + .5 * nc[i,3] - labcd$x[i] ) / xlen
1105     y <- ( nc[i,2] + .5 * nc[i,4] - labcd$y[i] ) / ylen
1106     dst <- sqrt( x^2 + y^2 )
1107     if ( dst > 0.05 ){
1108         segs <- rbind(
1109             segs,
1110             c(

```

```

1111                                     nc[i,1] + .5 * nc[i,3], nc[i,2] + .5 * nc[i
1112                                     ,4],
1113                                     xorg[i], yorg[i]
1114                                     )
1115                                 }
1116                            }
1117
1118                            xorg <- labcd$x
1119                            yorg <- labcd$y
1120                            labcd$x <- nc[,1] + .5 * nc[,3]
1121                            labcd$y <- nc[,2] + .5 * nc[,4]
1122                        }
1123
1124
1125    }
1126
1127    if ( exists("segs" ) ){
1128        if ( is.null(segs) == F ){
1129            for ( i in 1:nrow(segs) ){
1130                segments(
1131                    segs[i,1],segs[i,2],segs[i,3],segs[i,4],
1132                    col="gray60",
1133                    lwd=1
1134                )
1135            }
1136        }
1137    }
1138
1139    text(
1140        labcd$x,
1141        labcd$y,
1142        labels=word_labs,
1143        cex=cex,
1144        offset=0,
1145        font=text_font
1146    )
1147
1148    if ( exists("out_coord" ) == F ) {
1149        out_coord <- cbind(
1150            labcd$x / (n_nodes-0.5),
1151            labcd$y / (n_nodes-1)
1152        )
1153    }
1154
1155    points1<-head(points,n=190)
1156    par(new=T)
1157    plot(points1[,1],points1[,2],type="c",col="red")
1158
1159    points2<-tail(points,n=190)
1160    par(new=T)
1161    plot(points2[,1],points2[,2],type="c",col="blue")

```
