

明治大学総合数理学部

2019 年度

卒 業 研 究

ヘルスケアとウェアラブルのデータ等利活用技術に関する調査
研究

学位請求者 先端メディアサイエンス学科

清水 崇喜

目次

第 1 章	はじめに	2
第 2 章	ウェアラブル端末所有者と非所有者の分析	3
2.1	データの概要	3
2.2	データの加工	4
2.3	分析結果	6
2.4	t 検定	11
第 3 章	BMI 悪化の要因	12
3.1	ロジスティック回帰	12
第 4 章	おわりに	16
	参考文献	18
付録 A	ウェアラブル端末による行動推定	19
A.1	はじめに	19
A.2	基本定義	19
A.3	実験結果	22
A.4	おわりに	23

第 1 章

はじめに

現在多くの企業からウェアラブルデバイスが発売され、ウェアラブル端末を用いたデータ収集や、データ分析が行われてきている。ヘルスケア等分野においても同様であり、ウェアラブルデバイスの健康維持への利活用が期待される。

経済産業省によるウェアラブルデータやデータ活用による疾病・介護予防や次世代ヘルスケア報告書 [1] では、ウェアラブル端末を活用した健康指導の介入によって体重等の改善に効果があることが示されている。また日本電気株式会社 [2] や、みずほ情報総合研究所 [3] などの民間企業においても同様の利活用が進められている。厚生労働省による日本再興戦略 2016 [4] においても、レセプトや健康診断のデータにウェアラブル端末等の IoT によるデータ収集を活用すれば、よりリアルタイムで個人の状況に応じた、効果的なサービス提供が可能となることが報告されている。従って、ウェアラブル端末を利用している人は、利用していない人に比べて、健康に強い関心を持っていることが予想される。

そこで、ウェアラブルデータを保持している者と保持していない者で健康状態に差異があるか確認するために調査を行う。2013 年から 2018 年まで 6 年間の 8,000 名の検診データとウェアラブルデータの匿名加工情報を用いて、ウェアラブル端末の利用の有無による統計的な差を明らかにする。

第 2 章

ウェアラブル端末所有者と非所有者の分析

2.1 データの概要

本研究では、あるヘルスケアサービス事業者から提供された匿名加工済みヘルスケアデータから、表 2.1 に示す健康診断データ、ウェアラブル端末による血圧取得データ、医療費データを用いて行う。本研究で利用する健康診断データの属性を表 2.2、ウェアラブル端末による血圧データを表 2.3、レセプトデータを表 2.4 に示す。

表 2.1 データの各統計量

table	健康診断データ	血圧データ	医療費データ
属性数	49	6	10
期間	2008-2018	2017-2018	2004-2018
レコード数 [件]	964,636	120,886	19,350,822
対象者 [人]	198,741	4,038	288,623

表 2.2 健康診断データ

属性	変数名	型例
仮 id	receipt_ID	user_A
検診受診月	date	2016/05/01
BMI	bmi	23.2
収縮期血圧 [mmHg]	h_bp	104
拡張期血圧 [mmHg]	l_bp	64

表 2.3 ウェアラブル端末による血圧取得データ

属性	変数名	例
仮個人 id	receipt_ID	user_A
年月日	date	2017/11/01
収縮期血圧 [mmHg]	h_bp	132
拡張期血圧 [mmHg]	l_bp	95

表 2.4 レセプトデータ

属性	変数名	例
仮 id	receipt_ID	qpHbEdvyVi
年月日	date	2017/2/01
金額 [円]	cost	1,520

2.2 データの加工

ウェアラブルデータを保持している者と保持していない者で健康状態に差異があるか確認するために調査を行う。ウェアラブル端末で血圧データを取得しているユーザー 4,037 人、ウェアラブル端末による血圧データを所有していないユーザー 4,000 人を抽出し、この二つのグループ間において BMI、治療額等の差異が発生するか検討する。

健康診断データの BMI の統計量を表 2.5 に示す統計量において、ウェアラブル端末を利用しているユーザーの方が BMI 平均値、レセプト一件あたりの医療費平均ともにわずかに大きい。各グループの BMI の分布を図 2.1 に示す。斜線ありをウェアラブル端末所有者 (WR)、斜線なしをウェアラブル端末非所有者とする。分布においては、ウェアラブルを利用しているユーザー (WR) の BMI 平均値の方が利用していないユーザー (NoWR) に比べて BMI が多いことがわかる。したがって、ウェアラブルを利用しているユーザーの方が BMI が高くなる傾向が見られる。

表 2.5 健康診断データの BMI 統計量

BMI	ウェアラブル所有	ウェアラブル非所有
平均	23.386	22.904
最大値	47.2	42.06
最低値	14.5	14.5
標準偏差	3.618	3.763

表 2.6 医療費データの統計量

レセプト一件あたりの医療費	ウェアラブル所有	ウェアラブル非所有
平均	13,703.6	12,163.5
最大値	692,854	368053.3
最低値	2,058.1	730
標準偏差	20,244.2	18,058.3

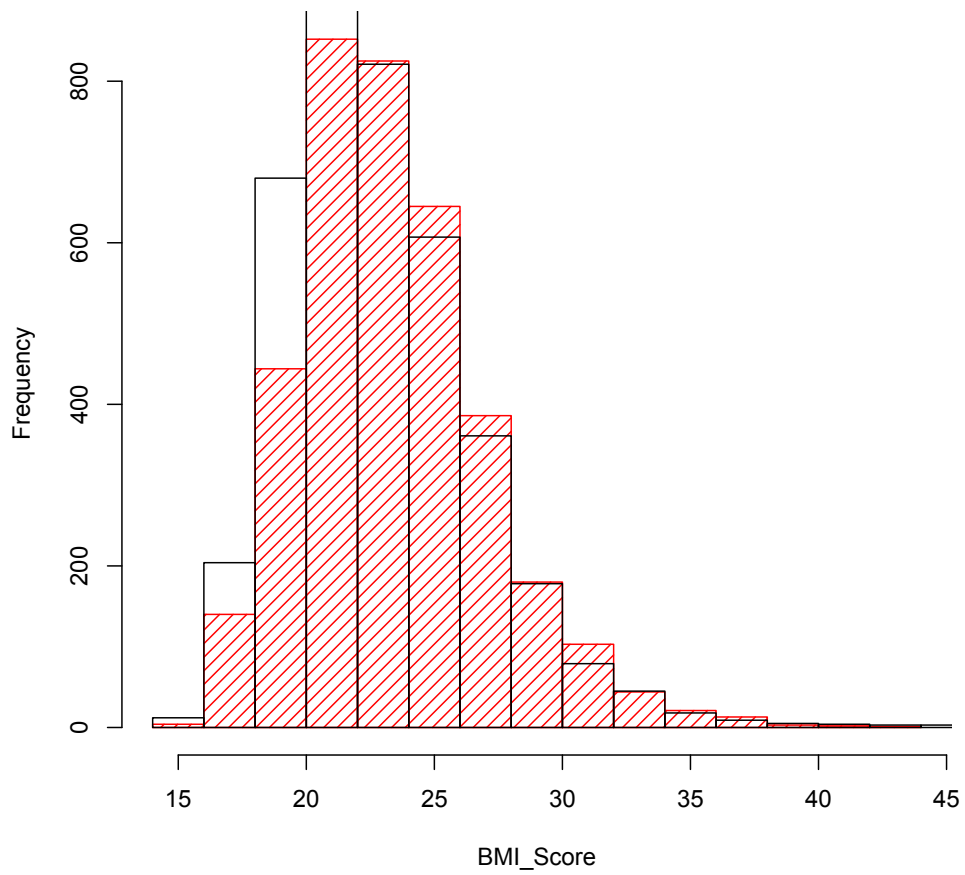


図 2.1 BMI 値の分布

表 2.7 BMI 評価 [7]

基準	区分	評価
18.5 未満	痩せ	U
18.5 以上 25 未満	普通	N
25 以上 30 未満	肥満 1	PO
30 以上 35 未満	肥満 2	O
35 以上 40 未満	肥満 3	O
40 以上	肥満 4	O

2.3 分析結果

ウェアラブル端末を使用しているグループと使用していないグループ間の BMI 数値を日本肥満学会の BMI 分類 [7] に従い、表 2.7 に示す 6 区分に分類して比較する。ウェアラブルを利用しているユーザにおける 1 回目の検診と 2 回目～6 回目の検診との BMI 指数の差を、それぞれ表 2.8～2.12 に示す。

ウェアラブルを利用していないユーザにおける 1 回目の検診と 2 回目～6 回目の検診との BMI 指数の差を、それぞれ表 2.18～2.17 に示す。

表 2.8 ウェアラブル所有者の 1 回目から 2 回目までの BMI 数値の推移

		2 回目						合計 [n]
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	
1 回目	痩せ	166	52	0	0	0	0	218
	普通	43	2,098	123	2	0	0	2,266
	肥満 1	0	101	686	40	1	0	828
	肥満 2	0	0	20	116	4	0	140
	肥満 3	0	0	1	1	19	1	22
	肥満 4	0	0	0	0	2	2	4
合計 [n]		209	2,365	882	159	26	3	3,478

表 2.9 ウェアラブル所有者の 1 回目から 3 回目までの BMI 数値の推移

		3 回目						合計 [n]
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	
1 回目	痩せ	155	52	0	0	0	0	207
	普通	33	1,979	139	0	0	0	2,151
	肥満 1	0	101	638	45	2	0	786
	肥満 2	0	1	31	97	6	0	135
	肥満 3	0	1	0	6	14	0	21
	肥満 4	0	0	0	1	2	1	4
合計 [n]		188	2,134	808	149	22	1	3,304

表 2.10 ウェアラブル所有者の 1 回目から 4 回目までの BMI 数値の推移

		4 回目						
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	合計 [n]
1 回目	痩せ	130	46	0	0	0	0	176
	普通	36	1,806	163	0	0	0	2,005
	肥満 1	0	123	580	46	0	2	751
	肥満 2	0	3	19	100	5	0	127
	肥満 3	0	0	0	4	12	0	16
	肥満 4	0	0	0	1	2	1	4
	合計 [n]	166	1,978	762	151	19	3	3,079

表 2.11 ウェアラブル所有者の 1 回目から 5 回目までの BMI 数値の推移

		5 回目						
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	合計 [n]
1 回目	痩せ	110	47	0	0	0	0	157
	普通	26	1,617	148	0	0	0	1,791
	肥満 1	0	126	506	51	2	0	685
	肥満 2	0	0	28	79	6	0	113
	肥満 3	0	0	0	1	11	0	12
	肥満 4	0	0	0	0	2	2	4
	合計 [n]	136	1,790	682	131	21	2	2,762

表 2.12 ウェアラブル所有者の 1 回目から 6 回目までの BMI 数値の推移

		6 回目						
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	合計 [n]
1 回目	痩せ	58	29	0	0	0	0	87
	普通	22	682	60	0	0	0	764
	肥満 1	0	60	187	27	1	0	275
	肥満 2	0	0	9	28	4	0	41
	肥満 3	0	0	0	1	3	0	4
	肥満 4	0	0	0	1	1	0	2
	合計 [n]	80	771	256	57	9	0	1,173

表 2.13 ウェアラブル非所有者の 1 回目から 2 回目までの BMI 数値の推移

		2 回目						
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	合計 [n]
1 回目	痩せ	229	56	0	0	0	0	285
	普通	57	2,007	85	0	0	0	2,149
	肥満 1	0	99	582	19	0	0	700
	肥満 2	0	1	17	98	5	0	121
	肥満 3	0	0	0	3	9	1	13
	肥満 4	0	0	1	1	0	5	7
	合計 [n]	286	2,163	684	122	14	6	3,275

表 2.14 ウェアラブル非所有者の 1 回目から 3 回目までの BMI 数値の推移

		2 回目						合計 [n]
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	
1 回目	痩せ	189	63	0	0	0	0	247
	普通	53	1,767	93	0	0	0	1,913
	肥満 1	0	108	512	27	0	0	647
	肥満 2	0	2	21	76	3	0	102
	肥満 3	0	0	0	1	7	1	9
	肥満 4	0	0	1	1	0	4	6
	合計 [n]	237	1,940	627	105	10	5	2,924

表 2.15 ウェアラブル非所有者の 1 回目から 4 回目までの BMI 数値の推移

		2 回目						合計 [n]
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	
1 回目	痩せ	139	72	0	0	0	0	211
	普通	45	1,493	134	0	0	0	1,672
	肥満 1	0	96	436	33	0	0	565
	肥満 2	0	1	19	61	5	0	86
	肥満 3	0	0	0	2	3		5
	肥満 4	0	0	1	1	2	2	6
	合計 [n]	184	1,662	590	97	10	2	2,545

表 2.16 ウェアラブル非所有者の 1 回目から 5 回目までの BMI 数値の推移

		2 回目						合計 [n]
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	
1 回目	痩せ	112	55	0	0	0	0	167
	普通	40	1,212	105	0	0	0	1,357
	肥満 1	0	81	353	34	0	0	468
	肥満 2	0	0	13	51	9	0	73
	肥満 3	0	0	0	0	2	1	3
	肥満 4	0	0	1	0	1	2	4
	合計 [n]	152	1,348	472	85	12	3	2,072

表 2.17 ウェアラブル非所有者の 1 回目から 6 回目までの BMI 数値の推移

		2 回目						合計 [n]
		痩せ	普通	肥満 1	肥満 2	肥満 3	肥満 4	
1 回目	痩せ	31	22	0	0	0	0	53
	普通	8	284	29	0	0	0	321
	肥満 1	0	25	97	11	0	0	133
	肥満 2	0	0	3	11	2	0	16
	肥満 3	0	0	0	0	2	0	2
	肥満 4	0	0	0	0	0	0	6
	合計 [n]	39	331	129	22	4	0	525

図 2.18 に BMI の変化を改善 (I), 維持 (R), 悪化 (D) に分類した推移図を示す. また 1 回目の健診から n 回目の検診時にウェアラブル所有ユーザ (WR) とウェアラブル非所有ユーザ (NoWR) でそれぞれ改善 (I) 維持 (R) 悪化 (D) を求めた図 2.2 とその変化率についてまとめた図 2.3 を示す. この二つの図より健康診断回数に応じた BMI の数値の変化は, ウェアラブルを利用しているユーザの方が長期にわたって安定している.

表 2.18 ウェアラブル端末所有者の 1 回目から 2 回目までの BMI 推移

		2 回目						合計
		改善 2	改善 1	維持	悪化 1	悪化 2	悪化 3	
1 回目	痩せ			166	52	0	0	218
	普通		43	2,098	123	2	0	2,266
	肥満 1	0	101	686	40	1	0	828
	肥満 2	0	20	116	4	0		140
	肥満 3	1	1	19	1			22
	肥満 4	0	2	2				4
	合計	1	167	3,087	220	3	0	3,478

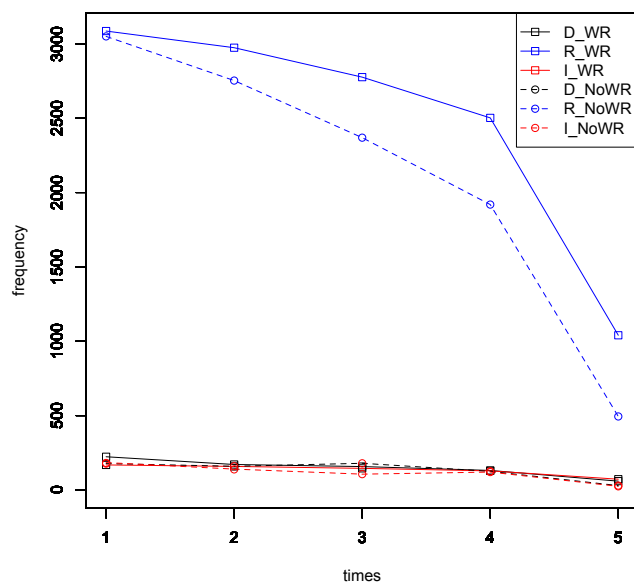


図 2.2 ウェアラブル端末と非所有者の BMI 数値の変化

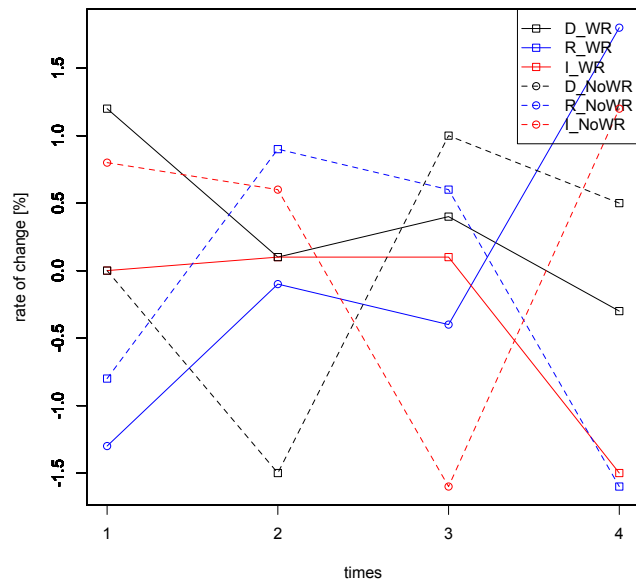


図 2.3 ウェアラブル端末所有者とウェアラブル端末非所有者の BMI 値の変化率

2.4 t 検定

表 2.5 と表 2.6 より、ウェアラブル端末を使用している被験者の方が BMI、医療費においてウェアラブル端末を使用していない被験者の数値を上回っていることがわかった。この二つのグループの差に統計的に有意であるか、前述の健診データのウェアラブル所有者 4,037 人とウェアラブル非所有者 4,000 人の BMI 値、医療費データを用いて有意水準 5 % で両側検定の等分散を仮定した 2 標本による t 検定を行った。表 2.19 に R の “t.test” で得られた BMI の検定結果を示す。表 2.20 に医療費の検定結果を示す。

表 2.19 BMI-t.test_result

t	-5.6878
df	7,626
p_value	1.335×10^{-8}

表 2.20 I-t.test_result

t	3.4614
df	7,538
p_value	0.0005404

検定の結果、医療費と BMI の P 値はそれぞれ $p=0.0005404 < .05$ 、 $p=1.335 \times 10^{-8} < .05$ であり、この二つのグループ間にある BMI の差は統計的に有意であると示された。

第 3 章

BMI 悪化の要因

3.1 ロジスティック回帰

ウェアラブル端末所有グループ内において、BMI の悪化にどのような要因が影響しているのか明らかにするため、式 (3.1) に示すロジスティック回帰分析を行った。

表 3.1 に示す様に、目的変数を「悪化 (D)」「維持 (R)」「改善 (I)」とし、10 変数を用いて分析した。表 3.2 に R の”glm”関数を用いて求めた目的変数が「悪化」の場合の結果、オッズ比を示す。表 3.3 に R の”glm”関数を用いて求めた目的変数が「維持」の場合の結果、オッズ比を示す。表 3.4 に R の”glm”関数を用いて求めた目的変数が「改善」の場合の結果、オッズ比を示す。

$$y = \frac{1}{1 + e^{-(a_1x_1 + a_2x_2 + a_ix_i)}} \quad (3.1)$$

表 3.1 分析に利用する説明変数

目的変数	説明変数
BMI の悪化	検診時収縮時血圧値平均 140 以上
	検診時拡張時血圧値平均 90 以上
	検診時 BMI25 以上
	ウェアラブル端末における血糖値データの有無
	ウェアラブル端末における血圧データの総数 300 以上
	ウェアラブル端末における歩行データの総数 300 以上
	ウェアラブル端末における歩行データの有無
	ウェアラブル端末で収集した 1 日平均歩数 8,000 歩以上

表 3.2 LogisticRegression Analtsys on WR

	Deterioration on WR
検診時収縮時血圧値 [Ave.]140 以上 (or)	0.258 (0.252)
検診時拡張時血圧値 [Ave.]90 以上	0.170 (0.278)
検診時 BMI25 以上	0.653*** (0.163)
ウェアラブル血糖値の有無	-11.311 (391.997)
ウェアラブル血圧データ数 300 以上	-0.523 (0.374)
ウェアラブル歩行データ数 300 以上	-0.269* (0.160)
ウェアラブル歩行データの有無	1.146 (1.017)
ウェアラブル歩行値 [Ave.]8,000 以上	-0.024 (0.211)
Constant	-4.301*** (1.013)
N	3,837
Log Likelihood	-684.619
AIC	1,387.238

***p < .01; **p < .05; *p < .1

ウェアラブルを利用しているユーザにおける健康診断 1 回目と 2 回目で BMI の数値が「悪化」する要因として「BMI が 25 以上」に有意差 $p=5.88 \times 10^{-5} < .05$ が見られた。よって BMI 評価において、肥満と判定される人は、次回の健康診断で、悪化する確率が高いことが示された。

表 3.3 LogisticRegression Analtsys on WR

	retain on WR
検診時収縮時血圧値 [Ave.]140 以上 (odds)	0.181 (0.155)
検診時拡張時血圧値 [Ave.]90 以上	-0.253 (0.163)
検診時 BMI25 以上	-0.147 (0.093)
ウェアラブル血糖値の有無	12.261 (237.784)
ウェアラブル血圧データ数 300 以上	0.383** (0.176)
ウェアラブル歩行データ数 300 以上	-0.121 (0.086)
ウェアラブル歩行データの有無	0.551* (0.301)
ウェアラブル歩行値 [Ave.]8,000 以上	-0.062 (0.108)
Constant	1.026*** (0.294)
N	3,837
Log Likelihood	-1,843.213
AIC	3,704.425

***p < .01; **p < .05; *p < .1

ウェアラブルを利用しているユーザにおける健康診断 1 回目と 2 回目で BMI の数値が「維持」する要因として、「ウェアラブル血圧データ数 300 以上」に有意差 $p=0.029247<.05$ 、「ウェアラブルの血圧データを 300 以上」に有意差 $p<0.043653$ が見られた。よって BMI 評価において、ユーザの BMI 値「維持」には日頃から継続的にデータを収集していることが要因に考えられる。

表 3.4 LogisticRegression Analtsys on WR

	Improvement on WR
検診時収縮時血圧値 [Ave.]140 以上 (odds)	0.051 (0.251)
検診時拡張時血圧値 [Ave.]90 以上	0.404 (0.256)
検診時 BMI25 以上	0.928*** (0.153)
ウェアラブル血糖値の有無	-11.786 (386.381)
ウェアラブル血圧データ数 300 以上	0.053 (0.258)
ウェアラブル歩行データ数 300 以上	0.324** (0.162)
ウェアラブル歩行データの有無	-0.614 (0.540)
ウェアラブル歩行値 [Ave.]8,000 以上	0.470*** (0.176)
Constant	-3.060*** (0.528)
N	3,837
Log Likelihood	-724.402
AIC	1,466.804

***p < .01; **p < .05; *p < .1

ウェアラブルを利用しているユーザにおける健康診断 1 回目と 2 回目で BMI の数値が「改善」する要因として、「BMI 数値 25 以上」に有意差 $p < 1.67e-09$, 「ウェアラブル歩行データ数 300 以上」に有意差 $p = 0.0075 < .05$, そして「一日平均歩数 8,000 歩以上」に有意差 $p < 0.00464$ が見られた。よって BMI 評価において、肥満と判定される人は、「悪化」だけではなく、「改善」にも働いていることが示された。また「維持」と同様に日常的にデータを取得しているユーザは、「維持」「改善」に効果があると考えられる。そして、1 日平均 8,000 歩以上歩いているユーザは、BMI の「改善」に大きな関連があると考えられる。

従って、表 3.2～表 3.4 の結果より、ウェアラブルを装着していることは、ある程度の健康改善に効果があると考えられる。

第 4 章

おわりに

本研究では、8,000 名のデータを使用して、ウェアラブルの利用しているユーザーとしていないユーザー、また利用しているユーザー間で、統計的にどのような差があらわれるのか、統計的に分析した。結果、ウェアラブルを装着していることは、ある程度の健康改善に効果があることが分かった。今後の課題として、ウェアラブルデータが 2017 年から 2018 年までと短く、健康診断データに比べてレコード数も少ないことから分析の精度を高めるために、継続的により中長期的にウェアラブルデータを採取していくことが必要である。

謝辞

本論文は筆者が明治大学総合数理学部先端メディアサイエンス学科学士課程に在籍中の研究成果をまとめたものである。本研究の完成は多くの方々からご指導とご援助が無ければ成し得なかった。ここに感謝の意を表す。特に学部2年時から3年間お世話になった明治大学総合数理学部先端メディアサイエンス学科教授、菊池浩明先生には指導教官として本研究の実施の機会を与えていただき、その遂行に当たって終始、ご指導をいただいた。また、学部3年時に添削、御助言等いただいた森駿文氏、四年生時匿名加工などをご指導、卒業論文を添削いただいた伊藤聡志氏に感謝の意を表す。更に、3年間苦楽を共にした明治大学菊池研究室の同期には、研究に対する有益な意見、プログラミングに対する様々な知見をいただいた。最後にここまで何不自由なく育ててくれた両親には、ここ明治大学で学ぶ機会をいただいた。本研究だけでなく、著者の学生生活は皆さまの支え無くしては成り立たなかった。この場を借りて、改めて深謝の意を表す。

参考文献

- [1] 経済産業省，“ウェアラブルデータやデータ活用による疾病・介護予防や次世代ヘルスケア”，平成 31 年 2 月
- [2] 日本電気株式会社，“医療等データの利活用技術に関する調査研究”，平成 28 年度臨床研究等 ICT 基盤構築研究事業
- [3] みずほ情報総研株式会社，“要介護認定等データ及び介護レセプトデータを用いた 要介護度変化の予測モデルにかかる実現可能性等の調査”，平成 30 年度 老人保健事業推進費等補助金 老人保健健康増進等事業
- [4] 厚生労働省，“日本再興戦略 2016”，厚生労働省
- [5] 21 世紀政策研究所，“データ利活用と産業化”，21 世紀政策研究所 研究プロジェクト，2018 年 5 月，
“www.21ppi.org/pdf/thesis/180531.pdf”
- [6] 一般社団法人日本経済団体連合会，“「Society 5.0 を実現するデータ活用推進戦略」目指すべきデータ活用社会のユースケース集”，2017 年 12 月，“keidanren.or.jp/policy/2017/104_usecase.pdf”
- [7] 日本肥満学会，“BMI 分類”，令和 2 年 1 月参照，“jasso.or.jp/data/magazine/pdf/chart_A.pdf”

付録 A

ウェアラブル端末による行動推定

A.1 はじめに

現在, 多くの分野において IoT の普及が始まっている. 多くの企業から IoT デバイスが発売され, データ分析も行われてきている.

本研究において, IoT デバイスの中でも代表的なものの 1 つであるウェアラブルデバイス, 特にスマートウォッチと呼ばれる腕時計型デバイスを用いてバイタルデータを取得する, 取得できるバイタルデータからその人物の行動を特定することはできるのかを, 推定をする.

A.2 基本定義

A.2.1 スマートウォッチについて

ウェアラブルデバイスの一種であり現在最も知られている IoT デバイスの一つである. 文字盤の部分がディスプレイになっているモノが多く, 内蔵のセンサなどを利用し, 生体情報などを計測することが出来る. 本実験における, スマートウォッチには主に Fitbit 社-Fitbit Versa, Huawei 社-Huawei Band 2 pro を利用し, これ以降 Fitbit Versa を Fitbit, Huawei Band2 pro を Huawei と呼ぶことにする.

A.2.2 アクティビティモードについて

それぞれの端末にはアクティビティモードが搭載されており機器の手動操作で持ってアクティビティモードを開始する. また一定時間の継続運動を機器が察知するとアクティビティとみなしてアクティビティモードに移行する. アクティビティモードでは, 詳細な生体情報と同時に, 携帯電話の GPS と接続することで連動して経路も記録することができる.

表 A.1 取得できるデータ，歩数の記録の例

	日付	時間	歩数
	2018-11-15	00:01:00	0
	2018-11-15	00:02:00	72
	2018-11-15	00:03:00	89

表 A.2 作成したデータベース sp2018.sqlite の概要

データベース	データ項目	形式
Steps	日付	yyyy-mm-dd
Distance	時間	hh:mm:ss
HeartRate	value	数値
Calories-Burn		

A.2.3 行動データベースの構築

Fitbit は計測したデータを Fitbit 社のサーバに保存し，登録したユーザーに格納データを提供する．本研究において，データをデータベースに格納するために python-Fitbit モジュールを使用し，Fitbit サーバから作成したデータベースに格納するプログラムを構築した．

A.2.4 行動データの作成

データの取得のために，被験者は 7 月 9 日から 11 月 24 日まで断続的にウェアラブルデバイスを手首に装着し，データを収集しデータベースに格納した．データ取得には，Fitbit 社-Fitbit Versa, Huawei 社-Huawei Band 2 pro を使用し，右手首尺骨に当たる部分に Fitbit, 左手同部分に Huawei を装着しデータを取得した．表 1 に Fitbit で収集したデータをデータベースに格納した行動の例を示す．表 1 では取得した単位時間あたりに記録した歩数を示している．表 2 に取得したデータの概要を示す．

A.2.5 データの収集

データを収集・分析するに当たって，予備実験として，自宅～中野キャンパス間を徒歩，自転車，自動車，電車＋徒歩の 4 方法でそれぞれ三往復し，Fitbit Versa, Huawei Band 2 pro のアクティビティ機能を使ってデータを計測，そのデータをまとめたアクティビティデータ T を作成した．表 3 に作成したアクティビティデータ T の概要を示す．データは，18 レコード 14 属性のデータである．アクティビティ機能には往路，復路の区別はないため，手動で入力した．「ペース」は 1 時間あたりに進む距離であり「アクティブタイム」は最も行動をしていた継続時間である．

A.2.6 箱ひげ図

データの特徴量を発見するため，それぞれの属性ごとに箱ひげ図を作成した．例として 4 状態のアクティビティにおける心拍数の箱ひげ図を載せる．

表 A.3 アクティビティデータ T

Fitbit	Style	データ件数	データ項目	項目
	by Walk	4	日付	yyyy/mm/dd
往復			往路 or 復路	
総経過時間			数値	
総距離			数値	
by cycle	4	平均速度	3 桁数値	
		最高速度	3 桁数値	
		最低速度	3 桁数値	
by car	6	平均心拍数	数値	
		最高心拍数	数値	
		最低心拍数	数値	
by train	4	消費カロリー	数値	
		総歩数	数値	
		アクティブタイム	数値	

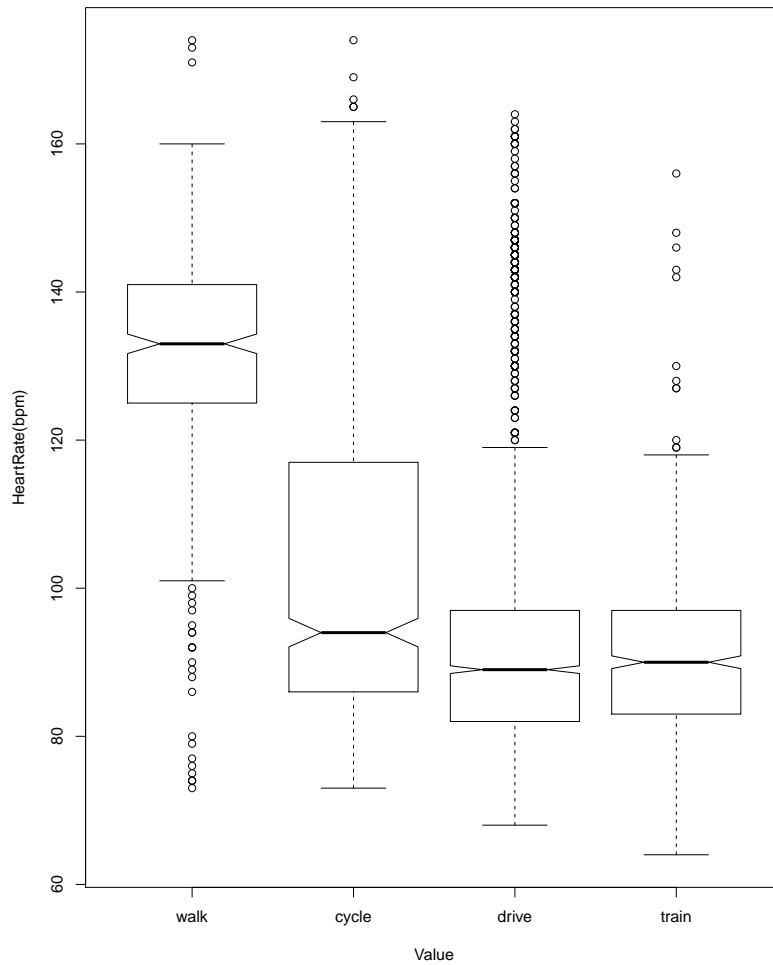


図 A.1 心拍数による箱ひげ図

A.2.7 決定木の作成

特徴量の判定のために、その該当データを R 言語の RPART によって決定木を作成した A.2. 決定木作成では目的変数に 4 状態、説明変数にデータ項目である日付、往復を除いた 12 項目を用いた。

A.3 実験結果

A.3.1 特徴量の算出

決定木から判定された、剪定に使われた説明変数を特徴量として決定した。

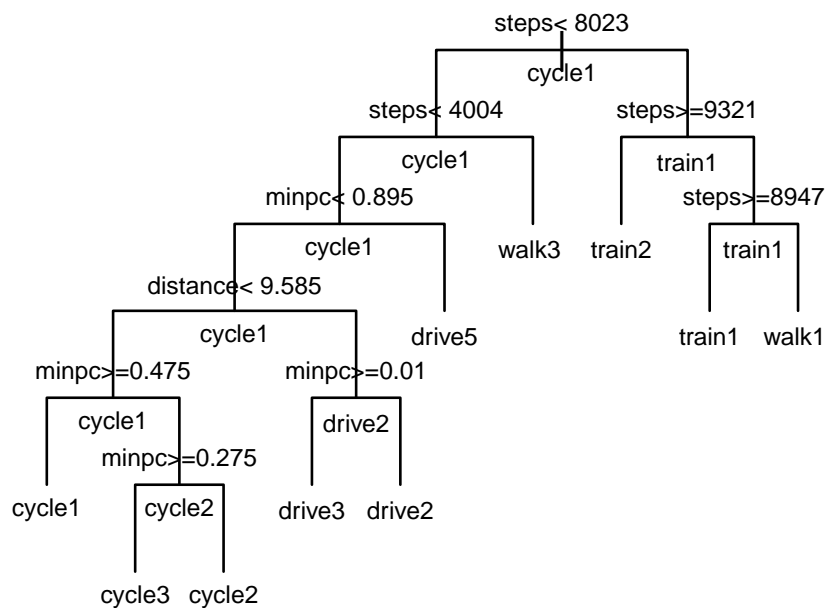


図 A.2 取得データによる決定木

A.3.2 決定木の判定結果

歩き 2 回目のデータのみ別のノードに含まれた。剪定には総歩数、最低速度、最高速度、距離が使用されていた。

A.3.3 考察

結果、RPART による決定木では、ノードの分岐に、総歩行数、最低速度、移動距離の三要素がしばしば用いられている。自転車、自動車による移動は徒歩や電車による移動に比べて歩行数が少ないため、この説明変数が判別に使用されたと考える。最低速度、最高速度に関しては、徒歩、自転車は時間が遅いながらも停止している時間が少ないため最低速度が 0 ではないが、電車、自動車は停止時間が発生するため、最低速度が使われ、また自動車や電車は、自転車、徒歩での移動に比べ最高時速が高いことから剪定に使われたと考える。距離について、電車、自動車が道を選びながら進まなければならないのに対して自転車、徒歩は比較的最短距離を選べるため使われたと考える

しかしながら RPART の説明変数記述する順番を変えると分岐に使用される説明変数が変わるため、総距離、最低速度、移動距離の三要素が 4 方法の推定に関する決定的な特徴量であるとは判別できない。

A.4 おわりに

本論では、データ数が 20 と少なく決定木で判定した特徴量に有用性があるかどうかは判定できなかった。データを収集する際に細かく自分の行動を明記し、そのデータに従って、判定できる特徴量を見つけるために研究を行なって行くことが必要である。