



UNIVERSITÀ DEGLI STUDI DI
MILANO-BICOCCA

F1801Q145

MODELLI PROBABILISTICI PER LE DECISIONI

HAR Bayesian Network

Studenti:

Artifoni Mattia

Brena Luca

Bottoni Federico

Matricole:

807466

808216

806944

Giugno 2019

Indice

1	Introduzione	2
1.1	Dominio di riferimento	2
1.2	Ipotesi e assunzioni	3
2	Scelte di design	3
2.1	Analisi statistica e qualitativa	3
2.2	Normalizzazione	3
2.3	Discretizzazione	3
3	I modelli di rete	3
3.1	pgmpy	3
3.2	Il modello correlato	4
4	Risultati e conclusioni	5

1 Introduzione

Il progetto ha l'obiettivo di creare un modello di Rete Bayesiana capace di predire il tipo di azione che sta effettuando un ipotetico individuo che indossa il "HAR wearable devices setup", una particolare sistema indossabile composto da 4 accelerometri che permette di analizzare i vettori accelerazione dei sensori in questione. Viene fornito dal progetto di riferimento[1] un dataset contenente dati sufficienti per effettuare training e testing del modello

1.1 Dominio di riferimento

La natura dei dati utilizzati è definita nel paper[3] del progetto di provenienza. La singola entry del dataset rappresenta uno snapshot acquisito dai sensori e consiste in:

- user: stringa rappresentante l'username dell'individuo in oggetto
- gender: stringa relativa al suo genere
- age: intero dell'età
- how_tall:in_meters: decimale che esprime l'altezza
- weight: intero riguardo al peso
- body_mass_index: decimale rappresentante la distribuzione del peso
- xi: intero che esprime la componente x del vettore accelerazione nel sensore i-esimo
- yi: intero che esprime la componente y del vettore accelerazione nel sensore i-esimo
- zi: intero che esprime la componente z del vettore accelerazione nel sensore i-esimo
- class: stringa, il target del modello sviluppato: rappresenta l'azione/situazione dell'individuo al momento dello snapshot, può assumere il valore di "walking", "standing", "standingup", "sitting" e "sittingdown"

1.2 Ipotesi e assunzioni

2 Scelte di design

2.1 Analisi statistica e qualitativa

user	gender	age	how_tall_in_meters	weight	body_mass_index	x1	y1	z1	x2	y2	z2	x3	y3	z3	x4	y4	z4	class
wallace	Man	31	1,71		83	28,4	-7	92	-114	19	89	-121	26	124	-83	-163	-76	-156 standing
katia	Woman	28	1,58		55	22,0	-13	43	-14	-12	-21	-21	15	104	-70	-141	-95	-174 sitting
debora	Woman	46	1,62		75	28,6	-11	90	-96	-5	70	-124	10	103	-86	-159	-104	-162 standing
wallace	Man	31	1,71		83	28,4	6	97	-119	-480	-502	-603	14	126	-98	-235	-89	-194 walking
katia	Woman	28	1,58		55	22,0	-8	92	-116	-493	-516	-616	34	11	-124	-187	-78	-157 walking
katia	Woman	28	1,58		55	22,0	-2	131	-131	-492	-515	-614	13	97	-46	-221	-101	-142 standingup
debora	Woman	46	1,62		75	28,6	-12	94	-99	0	72	-124	14	103	-92	-170	-104	-162 standing
debora	Woman	46	1,62		75	28,6	1	90	-48	-21	23	-17	8	108	-88	-164	-94	-155 sitting
debora	Woman	46	1,62		75	28,6	-1	89	-48	-18	23	-16	15	108	-90	-163	-93	-157 sitting
katia	Woman	28	1,58		55	22,0	17	138	-148	-492	-515	-615	23	149	-122	-165	-56	-176 walking
debora	Woman	46	1,62		75	28,6	-2	90	-50	-20	21	-19	14	107	-92	-158	-95	-160 sitting
katia	Woman	28	1,58		55	22,0	-1	102	-96	16	88	-121	44	103	-91	-203	-87	-163 standing
jose_carlos	Man	75	1,67		67	24,0	-12	96	-101	-11	-19	-21	-65	88	-104	-100	-143	-136 sitting
debora	Woman	46	1,62		75	28,6	0	106	-84	-5	86	-123	32	106	-84	-177	-89	-156 standing
wallace	Man	31	1,71		83	28,4	-11	95	-165	35	74	-131	31	180	-57	-238	-63	-161 walking
wallace	Man	31	1,71		83	28,4	4	71	-61	-22	-25	-21	66	58	-101	-93	-126	-164 sitting
katia	Woman	28	1,58		55	22,0	-1	101	-88	-5	87	-124	25	92	-49	-199	-90	-159 standing
katia	Woman	28	1,58		55	22,0	-29	26	-21	-261	-372	-452	27	109	-86	-150	-92	-169 sitting
wallace	Man	31	1,71		83	28,4	26	60	-158	-489	-512	-612	-55	167	-81	-220	-41	-169 walking
katia	Woman	28	1,58		55	22,0	1	103	-95	-5	83	-127	42	101	-61	-185	-81	-155 standing
debora	Woman	46	1,62		75	28,6	-7	85	-86	-74	21	-163	45	54	-143	-186	-95	-159 walking
wallace	Man	31	1,71		83	28,4	-20	155	-131	-5	154	-41	-429	176	-112	-192	-81	-162 walking
wallace	Man	31	1,71		83	28,4	3	64	-59	-37	-24	-27	74	68	-106	-81	-141	-157 sitting
jose_carlos	Man	75	1,67		67	24,0	-14	95	-97	-8	-10	-20	-63	86	-128	-101	-148	-131 sitting
debora	Woman	46	1,62		75	28,6	-5	74	-36	-4	20	-33	4	96	-69	-151	-107	-164 sitting
wallace	Man	31	1,71		83	28,4	6	72	-58	-33	-4	-23	66	59	-103	-98	-122	-183 sitting
wallace	Man	31	1,71		83	28,4	14	97	-120	-490	-513	-613	22	123	-94	-209	-93	-176 walking
wallace	Man	31	1,71		83	28,4	4	66	-59	-37	-24	-27	69	62	-110	-87	-145	-166 sitting
katia	Woman	28	1,58		55	22,0	-10	91	-159	2	31	-40	18	104	-56	-269	-43	-121 standingup
katia	Woman	28	1,58		55	22,0	-13	39	-15	-8	-23	-23	29	107	-80	-149	-95	-165 sitting
wallace	Man	31	1,71		83	28,4	26	98	-84	-20	9	-27	75	68	-61	-106	-83	-148 standingup

Figura 1: Dataset dopo lo shuffle

user	gender	age	how_tall_in_meters	weight	body_m	x1	y1	z1	x2	y2	z2	x3	y3	z3	x4	y4	z4	class
122039	jose_carlos	Man	75	1,67	67	24,0	-8	95	-141	-11	71	-58	-17	111	-100	-214	-95	-154 standingup
122042	jose_carlos	Man	75	1,67	67	24,0	-8	95	-114	-18	87	-98	12	122	-105	-199	-84	-146 standingup
122059	jose_carlos	Man	75	1,67	67	24,0	-8	94	-115	-18	87	-91	10	122	-104	-195	-84	-142 standingup
122069	jose_carlos	Man	75	1,67	67	24,0	-8	93	-145	-18	68	-58	-9	120	-97	-219	-98	-148 standingup
122078	jose_carlos	Man	75	1,67	67	24,0	-8	101	-120	-13	91	-101	17	123	-108	-207	-82	-14420-11-2011 04:50:23.713 standingup
122079	jose_carlos	Man	75	1,67	67	24,0	-8	93	-104	7	-40	-24	-16	116	-86	-247	-139	-104 standingup
122080	jose_carlos	Man	75	1,67	67	24,0	-8	97	-103	6	-41	-24	-17	113	-87	-301	-140	-88 standingup
122103	jose_carlos	Man	75	1,67	67	24,0	-8	89	-167	-4	25	-29	-24	113	-83	-273	-81	-145 standingup
122121	jose_carlos	Man	75	1,67	67	24,0	-8	97	-145	7	-9	-22	-18	116	-92	-247	-102	-162 standingup

Figura 2: Errore in un entry del dataset

2.2 Normalizzazione

2.3 Discretizzazione

3 I modelli di rete

3.1 pgmpy

Il software scelto è pgmpy[2] di Python, una libreria che permette di modellare le dipendenze in modo agile, stimare le CPT delle variabili sfruttando dei metodi che ac-

cettano il dataset ed effettuare inferenze dichiarando la variabile di query e le evidenze.

Utilizzando la libreria ci siamo resi conto di come sia performante utilizzando modelli semplici e correlati da pochi record, tuttavia appena è avvenuta l'esecuzione della stima delle CPT nel primo modello completo ideato abbiamo riscontrato le prime difficoltà: il modello decisamente complesso viene bloccato da Python date le eccessive combinazioni possibili. Abbiamo quindi tentato di semplificare la rete.

3.2 Il modello correlato

x1	y1	0.345808064	x2	x1	0.198233156	x3	x1	0.07121	x4	x1	0.099018401
x1	z1	0.030417499	x2	y1	0.043861069	x3	y1	-0.1816	x4	y1	-0.29045831
x1	x2	0.198233156	x2	z1	0.109970588	x3	z1	0.13028	x4	z1	0.50628732
x1	y2	0.202451431	x2	y2	0.977115735	x3	x2	0.00802	x4	x2	0.157007278
x1	z2	0.251138704	x2	z2	0.953083727	x3	y2	0.00636	x4	y2	0.081715165
x1	x3	0.071207709	x2	x3	0.008022688	x3	z2	0.00665	x4	z2	0.287828127
x1	y3	-0.136539934	x2	y3	-0.140455652	x3	y3	0.32843	x4	x3	0.166694382
x1	z3	0.00417011	x2	z3	0.107520027	x3	z3	0.27106	x4	y3	-0.111224183
x1	x4	0.099018401	x2	x4	0.157007278	x3	x4	0.16669	x4	z3	0.035357614
x1	y4	-0.142551934	x2	y4	-0.23983578	x3	y4	0.04272	x4	y4	-0.600982199
x1	z4	-0.025592835	x2	z4	0.164505939	x3	z4	-0.2023	x4	z4	-0.068008246
y1	x1	0.345808064	y2	x1	0.202451431	y3	x1	-0.1365	y4	x1	-0.142551934
y1	z1	-0.5159614	y2	y1	0.138219068	y3	y1	0.19162	y4	y1	0.228997237
y1	x2	0.043861069	y2	z1	0.017749163	y3	z1	-0.119	y4	z1	-0.405502292
y1	y2	0.138219068	y2	x2	0.977115735	y3	x2	-0.1405	y4	x2	-0.23983578
y1	z2	-0.0301789	y2	z2	0.918648041	y3	y2	-0.096	y4	y2	-0.15437806
y1	x3	-0.181573944	y2	x3	0.006358901	y3	z2	-0.2002	y4	z2	-0.389086436
y1	y3	0.191617913	y2	y3	-0.095987852	y3	x3	0.32843	y4	x3	0.042718189
y1	z3	0.109626835	y2	z3	0.120384491	y3	z3	0.67093	y4	y3	0.3239336
y1	x4	-0.29045831	y2	x4	0.081715165	y3	x4	-0.1112	y4	z3	0.076057385
y1	y4	0.228997237	y2	y4	-0.15437806	y3	y4	0.32393	y4	x4	-0.600982199
y1	z4	0.186791537	y2	z4	0.165709979	y3	z4	-0.0364	y4	z4	-0.117144404
z1	x1	0.030417499	z2	x1	0.251138704	z3	x1	0.00417	z4	x1	-0.025592835
z1	y1	-0.5159614	z2	y1	-0.0301789	z3	y1	0.10963	z4	y1	0.186791537
z1	x2	0.109970588	z2	z1	0.2172898	z3	z1	0.12427	z4	z1	-0.197813141
z1	y2	0.017749163	z2	x2	0.953083727	z3	x2	0.10752	z4	x2	0.164505939
z1	z2	0.2172898	z2	y2	0.918648041	z3	y2	0.12038	z4	y2	0.165709979
z1	x3	0.130282473	z2	x3	0.006650803	z3	z2	0.07906	z4	z2	0.160719903
z1	y3	-0.118961498	z2	y3	-0.200237814	z3	x3	0.27106	z4	x3	-0.202266103
z1	z3	0.124272757	z2	z3	0.07905866	z3	y3	0.67093	z4	y3	-0.036427754
z1	x4	0.50628732	z2	x4	0.287828127	z3	x4	0.03536	z4	z3	0.031079739
z1	y4	-0.405502292	z2	y4	-0.389086436	z3	y4	0.07606	z4	x4	-0.068008246
z1	z4	-0.197813141	z2	z4	0.160719903	z3	z4	0.03108	z4	y4	-0.117144404

Figura 3: Indice di correlazione di Paerson calcolato su tutte le combinazioni di componenti

4 Risultati e conclusioni

Riferimenti bibliografici

- [1] <http://groupware.les.inf.puc-rio.br/har>.
- [2] <http://pgmpy.org/>.
- [3] Katia Vega Eduardo Velloso Ruy Milidiú Wallace Ugulino, Débora Cardador and Hugo Fuks. Wearable computing: Accelerometers' data classification of body postures and movements, 2012.