2022-04-29

Leo Clo, Alejandro Hohmann, Garrett Michael (on behalf of Spotify)

Project: *WISDM Smartphone and Smartwatch Activity and Biometrics Dataset*

- **Dataset description (2)**
  - Accelerometer and Gyroscope data take from Phone and Smartwatch in 51 different subjects performing 18 different tasks. Measures movement data over ten-second intervals while subjects perform the various tasks.

TABLE 1
SUMMARY INFORMATION FOR THE DATASETS

Number of subjects	51
Number of activities	18
Minutes collected per activity	3
Sensor polling rate	20Hz
Smartphone used	Google Nexus 5/5x or Samsung Galaxy S5
Smartwatch used	LG G Watch
Number raw measurements	15,630,426

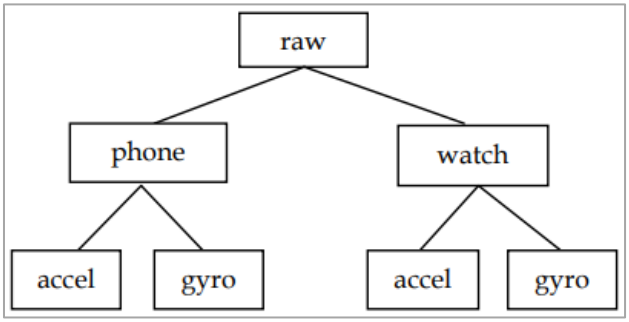

TABLE 2
THE 18 ACTIVITIES REPRESENTED IN DATA SET

Activity	Code
Walking	A
Jogging	B
Stairs	C
Sitting	D
Standing	E
Typing	F
Brushing Teeth	G
Eating Soup	H
Eating Chips	I
Eating Pasta	J
Drinking from Cup	K
Eating Sandwich	L
Kicking (Soccer Ball)	M
Playing Catch w/Tennis Ball	O
Dribblinlg (Basketball)	P
Writing	Q
Clapping	R
Folding Clothes	S

TABLE 4
DISTRIBUTION OF RAW SENSOR DATA

Activity	Phone		Watch		Total	Class %
	Accel	Gyro	Accel	Gyro		
Walking	279,817	203,919	210,495	192,531	886,762	5.7%
Jogging	268,409	200,252	205,787	187,833	862,281	5.5%
Stairs	255,645	197,857	207,312	180,416	841,230	5.4%
Sitting	264,592	202,370	213,018	195,050	875,030	5.6%
Standing	269,604	202,351	216,529	194,103	882,587	5.6%
Typing	246,356	194,540	205,137	187,175	833,208	5.3%
Brush Teeth	269,609	202,622	208,720	190,759	871,710	5.6%
Eat Soup	270,756	202,408	209,483	187,057	869,704	5.6%
Eat Chips	261,360	197,905	210,048	192,085	861,398	5.5%
Eat Pasta	249,793	197,844	203,112	189,609	840,358	5.4%
Drinking	285,190	202,395	215,879	197,917	901,381	5.8%
Eat Sandwich	265,781	197,915	203,684	190,191	857,571	5.5%
Kicking	278,766	202,625	209,491	191,535	882,417	5.6%
Catch	272,219	198,756	210,107	187,684	868,766	5.6%
Dribblinlg	272,730	202,331	212,810	194,845	882,716	5.6%
Writing	260,497	197,894	215,365	197,403	871,159	5.6%
Clapping	268,065	202,330	208,734	190,776	869,905	5.6%
Fold Clothes	265,214	202,321	211,335	193,373	872,243	5.6%
Total	4,804,403	3,608,635	3,777,046	3,440,342	15,630,426	100%

- **Analysis task to perform (2)**
  - Wrangle data: 15M+ records across 200+ files
  - Create predictive *classification* model(s) of what type of task subject is engaged in

- **Problem definition (2)**
    - Dynamically target advertising based on listeners activity
        - e.g. while user is listening to a podcast and folding their laundry, serve them an ad for laundry detergent
    - Match type of music recommendation / auto-play to the type of activity
        - e.g. while use is exercise, play up-tempo music
        - e.g. while user is eating pasta, play romantic music
- **Insights to be gained (2)**
    - Does the measurement data for one user performing an activity match the measurement data of another user (or do they require different fits per user?)
    - Do the smartphone and smartwatch require different predictive models?
- **Potential challenges with data and/or task (1)**
    - Really large data set (15M+ records)
        - Understanding how to map the measurement files to the subject and their activity
    - Working with pre-aggregated time series data which is aggregated to ten second intervals
        - Current aggregation may not be suitable for our modeling. Data set also contains the raw time series data which we may need to transform for our modelling purposes.
- **All team members present (1)**
    - We are 😊