



RAPIDS Framework

“Reproducible Analysis Pipeline for Data Streams (RAPIDS) allows you to process smartphone and wearable data to **extract** and **create behavioral features** (a.k.a. digital biomarkers), **visualize** mobile sensor data, and **structure** your analysis into reproducible workflows.”

How to get features from AWARE to RAPIDS

Download AWARE Data in CSV format

Data is downloaded off AWARE in a format called "aware_csv" (datastream compatible with RAPIDS; has the same column names)

Run through RAPIDS

Edit the config.yml in the RAPIDS repo + add a .yml for each participant

Sample .ymls available in repo for project

Obtain CSV of processed features

Processed features are saved in the RAPIDS directory under "processed"

Example of processed data: call logs

AWARE Raw Data

_id	timestamp	device_id	call_type	call_duration	trace
1	1.58766E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		2	14 d5e84f8af01b2728021d4f43f53a163c0c90000c
2	1.58774E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		3	0 47c125dc7bd163b8612cdea13724a814917b6e93
5	1.58775E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		2	95 9cc793ffd6e88b1d850ce540b5d7e000ef5650d4
6	1.58791E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		2	63 51fb9344e988049a3fec774c7ca622358bf80264
7	1.58799E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		3	0 2a862a7730cfdaf103a9487afe3e02935fd6e02
8	1.58802E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		1	11 a2c53f6a086d98622c06107780980cf1bb4e37bd
11	1.58818E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		2	65 56589df8c830c70e330b644921ed38e08d8fd1f3
12	1.5882E+12	a748ee1a-1d0b-4ae9-9074-279a2b6ba524		3	0 cab458018a8ed3b626515e794c70b6f415318adc
13	1.5882E+12	b748ee1a-1d0b-4ae9-9074-279a2b6ba524		3	0 zab458018a8ed3b626515e794c70b6f415318adc

Features to extract

local_segment local_segment_label,
local_segment_end_datetime,
phone_calls_rapids_missed_distinctcontacts
phone_calls_rapids_missed_timelastcall
phone_calls_rapids_incoming_count
phone_calls_rapids_incoming_meanduration
phone_calls_rapids_incoming_minduration
phone_calls_rapids_incoming_stdduration
phone_calls_rapids_incoming_entropyduration
phone_calls_rapids_incoming_timelastcall
phone_calls_rapids_outgoing_count
phone_calls_rapids_outgoing_meanduration
phone_calls_rapids_outgoing_minduration
phone_calls_rapids_outgoing_stdduration
phone_calls_rapids_outgoing_entropyduration
phone_calls_rapids_outgoing_timelastcall
pid (patient ID)

local_segment_start_datetime
phone_calls_rapids_missed_count
phone_calls_rapids_missed_timefirstcall
phone_calls_rapids_missed_countmostfrequentcontact
phone_calls_rapids_incoming_distinctcontacts
phone_calls_rapids_incoming_sumduration
phone_calls_rapids_incoming_maxduration
phone_calls_rapids_incoming_modeduration
phone_calls_rapids_incoming_timefirstcall
phone_calls_rapids_incoming_countmostfrequentcontact
phone_calls_rapids_outgoing_distinctcontacts
phone_calls_rapids_outgoing_sumduration
phone_calls_rapids_outgoing_maxduration
phone_calls_rapids_outgoing_modeduration
phone_calls_rapids_outgoing_timefirstcall
phone_calls_rapids_outgoing_countmostfrequentcontact,

Processed Data

	local_segme	local_segment	local_segment	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls
04-21	afternoon	4/23/20 12:00	4/23/20 17:59	0	0	NA	NA	0	0	0	NA	NA	NA	NA	NA
04-21	afternoon	4/24/20 12:00	4/24/20 17:59	0	0	NA	NA	0	0	0	NA	NA	NA	NA	NA
04-21	afternoon	4/27/20 12:00	4/27/20 17:59	0	0	NA	NA	0	1	1	11	11	11	11	NA
04-21	afternoon	4/29/20 12:00	4/29/20 17:59	0	0	NA	NA	0	0	0	NA	NA	NA	NA	NA
04-23:5	daily	4/23/20 0:00	4/23/20 23:59	0	0	NA	NA	0	0	0	NA	NA	NA	NA	NA
04-23:5	daily	4/24/20 0:00	4/24/20 23:59	1	1	638	638	1	0	0	NA	NA	NA	NA	NA
04-23:5	daily	4/26/20 0:00	4/26/20 23:59	0	0	NA	NA	0	0	0	NA	NA	NA	NA	NA
04-23:5	daily	4/27/20 0:00	4/27/20 23:59	1	1	544	544	0	1	1	11	11	11	11	NA
04-23:5	daily	4/29/20 0:00	4/29/20 23:59	1	1	1082	1082	0	0	0	NA	NA	NA	NA	NA
04-29 2	evening	4/29/20 18:00	4/29/20 23:59	1	1	1082	1082	1	0	0	NA	NA	NA	NA	NA
04-24	morning	4/24/20 6:00	4/24/20 11:59	1	1	638	638	1	0	0	NA	NA	NA	NA	NA
04-26	morning	4/26/20 6:00	4/26/20 11:59	0	0	NA	NA	0	0	0	NA	NA	NA	NA	NA
04-27	morning	4/27/20 6:00	4/27/20 11:59	1	1	544	544	0	0	0	NA	NA	NA	NA	NA
04-25	threeday	4/23/20 0:00	4/25/20 23:59	1	1	638	638	1	0	0	NA	NA	NA	NA	NA
04-26	threeday	4/24/20 0:00	4/26/20 23:59	1	1	638	638	1	0	0	NA	NA	NA	NA	NA
04-27	threeday	4/25/20 0:00	4/27/20 23:59	1	1	544	544	1	1	1	11	11	11	11	NA
04-28	threeday	4/26/20 0:00	4/28/20 23:59	1	1	544	544	0	1	1	11	11	11	11	NA

phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	phone_calls	pid
NA	NA	NA	NA	0	1	1	14	14	14	14	NA	14	0	814	814	1	p01
NA	NA	NA	NA	0	1	1	95	95	95	95	NA	95	0	762	762	0	p01
11	0	1000	1000	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	p01
NA	NA	NA	NA	0	1	1	65	65	65	65	NA	65	0	723	723	0	p01
NA	NA	NA	NA	0	1	1	14	14	14	14	NA	14	0	814	814	1	p01
NA	NA	NA	NA	0	1	1	95	95	95	95	NA	95	0	762	762	0	p01
NA	NA	NA	NA	0	1	1	63	63	63	63	NA	63	0	629	629	0	p01
11	0	1000	1000	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	p01
NA	NA	NA	NA	0	1	1	65	65	65	65	NA	65	0	723	723	0	p01
NA	NA	NA	NA	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	p01
NA	NA	NA	NA	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	p01
NA	NA	NA	NA	0	1	1	63	63	63	63	NA	63	0	629	629	1	p01
NA	NA	NA	NA	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	p01
NA	NA	NA	NA	0	2	2	54.5	109	14	95	57.2756493	14	0.38799827	814	762	1	p01
NA	NA	NA	NA	0	2	2	79	158	63	95	22.627417	95	0.67565964	762	629	1	p01



Features used in other papers, eg:

Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data

Sensor	Data collected
Battery	Initiation of phone charging, unplugging of device, battery level
ESM	
Locations	Participant's current location based on GPS latitude and longitude coordinates
Screen/ Device Usage plugin	Phone usage and non-usage sessions measured using the screen sensor. A usage session is defined as the time from when phone is unlocked until it is locked. Non-usage is defined as the time from when phone is locked until it is unlocked.
Timezone	User's current timezone based on GPS at the time of measurement

Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence

1. Heart rate: Mean, median, number of minutes in fat burn zone
2. Sleep: Average sleep, sum of minutes in bed after waking up from nap, sum of minutes in bed awake
3. Step: Number of active + sedentary segments in one day
4. Activity: The total duration of on foot, running, and on bicycle activities
5. Battery: Total duration of charging + discharging segments
6. Call: Shannon entropy for the duration of incoming + outgoing calls
7. Audio: The standard deviation of the level of noise in one day, the ratio between noise and whole conversation (silence, noise, voice, unknown), the minimum of the level of voice in one day
8. Locations: The maximum distance from home in meters, the number of significant locations visited in one day, the standard deviation of the length of all movements, the estimate of Shannon entropy for the significant locations visited in one day, the shortest time the user staying at a significant location, the ratio between the time of moving status and static status
9. Screen: The number of all unlock segments, hours since midnight of the first unlock in one day, length of the shortest unlock segment

Some of these features may require FitBit data

Common features across papers:

1. Battery charging/discharging duration (info available via AWARE)
2. Locations - distance from home, radius from home, variety of locations (info available via AWARE)
3. Amount of time unlocked/ phone usage with calling duration and amount of texting (info available via AWARE)
4. Time in bed/ time spent sleeping (info available via ESM)



Data streams & features...

Features available for Android and iOS unless otherwise specified



Battery

AWARE Raw Data

RAPIDS column	Stream column
TIMESTAMP	timestamp
DEVICE_ID	device_id
BATTERY_STATUS	battery_status
BATTERY_LEVEL	battery_level
BATTERY_SCALE	battery_scale

Processed Data

Feature	Units	Description
countdischarge	episodes	Number of discharging episodes.
sumdurationdischarge	minutes	The total duration of all discharging episodes.
countcharge	episodes	Number of battery charging episodes.
sumdurationcharge	minutes	The total duration of all charging episodes.
avgconsumptionrate	episodes/minutes	The average of all episodes' consumption rates. An episode's consumption rate is defined as the ratio between its battery delta and duration
maxconsumptionrate	episodes/minutes	The highest of all episodes' consumption rates. An episode's consumption rate is defined as the ratio between its battery delta and duration

Phone Screen

AWARE Raw Data

RAPIDS column	Stream column
TIMESTAMP	timestamp
DEVICE_ID	device_id
SCREEN_STATUS	screen_status

Processed Data

Feature	Units	Description
sumduration	minutes	Total duration of all unlock episodes.
maxduration	minutes	Longest duration of any unlock episode.
minduration	minutes	Shortest duration of any unlock episode.
avgduration	minutes	Average duration of all unlock episodes.
stdduration	minutes	Standard deviation duration of all unlock episodes.
countepisode	episodes	Number of all unlock episodes
firstuseafter	minutes	Minutes until the first unlock episode.

Phone conversations - **Android only**

Processed Data

AWARE Raw Data

RAPIDS column	Stream column
TIMESTAMP	timestamp
DEVICE_ID	device_id
DOUBLE_ENERGY	double_energy
INFERENCE	inference
DOUBLE_CONVO_START	double_convo_start
DOUBLE_CONVO_END	double_convo_end

Feature	Units	Description			
			noiseavgenergy	L2-norm	Average of all energy values when inference is noise
minutessilence	minutes	Minutes labeled as silence			
			noisesdenergy	L2-norm	Standard Deviation of all energy values when inference is noise
minutesnoise	minutes	Minutes labeled as noise			
			noiseminenergy	L2-norm	Minimum of all energy values when inference is noise
minutesvoice	minutes	Minutes labeled as voice			
			noisemaxenergy	L2-norm	Maximum of all energy values when inference is noise
minutesunknown	minutes	Minutes labeled as unknown			
sumconversationduration	minutes	Total duration of all conversations			
			voicesumenergy	L2-norm	Sum of all energy values when inference is voice
maxconversationduration	minutes	Longest duration of all conversations			
			voiceavgenergy	L2-norm	Average of all energy values when inference is voice
minconversationduration	minutes	Shortest duration of all conversations			
			voicesdenergy	L2-norm	Standard Deviation of all energy values when inference is voice
avgconversationduration	minutes	Average duration of all conversations			
			voiceminenergy	L2-norm	Minimum of all energy values when inference is voice
sdconversationduration	minutes	Standard Deviation of the duration of all conversations			
			voicemaxenergy	L2-norm	Maximum of all energy values when inference is voice
timefirstconversation	minutes	Minutes since midnight when the first conversation for a time segment was detected			
			silencesensedfraction	-	Ratio between minutessilence and the sum of (minutessilence, minutesnoise, minutesvoice, minutesunknown)
timelastconversation	minutes	Minutes since midnight when the last conversation for a time segment was detected			
			noiseexpectedfraction	-	Ration between minutesnoise and the number of minutes that in theory should have been sensed based on the record and pause cycle of the plugin (1440 / recordingMinutes+pausedMinutes)
noisesumenergy	L2-norm	Sum of all energy values when inference is noise			
			voiceexpectedfraction	-	Ration between minutesvoice and the number of minutes that in theory should have been sensed based on the record and pause cycle of the plugin (1440 / recordingMinutes+pausedMinutes)
noisesensedfraction	-	Ratio between minutesnoise and the sum of (minutessilence, minutesnoise, minutesvoice, minutesunknown)			
voicesensedfraction	-	Ratio between minutesvoice and the sum of (minutessilence, minutesnoise, minutesvoice, minutesunknown)			
unknownsensedfraction	-	Ratio between minutesunknown and the sum of (minutessilence, minutesnoise, minutesvoice, minutesunknown)			
			unknownexpectedfraction	-	Ration between minutesunknown and the number of minutes that in theory should have been sensed based on the record and pause cycle of the plugin (1440 / recordingMinutes+pausedMinutes)
silenceexpectedfraction	-	Ration between minutessilence and the number of minutes that in theory should have been sensed based on the record and pause cycle of the plugin (1440 / recordingMinutes+pausedMinutes)			

Locations

AWARE Raw Data

RAPIDS column	Stream column
TIMESTAMP	timestamp
DEVICE_ID	device_id
DOUBLE_LATITUDE	double_latitude
DOUBLE_LONGITUDE	double_longitude
DOUBLE_BEARING	double_bearing
DOUBLE_SPEED	double_speed
DOUBLE_ALTITUDE	double_altitude
PROVIDER	provider
ACCURACY	accuracy

Processed Data

Feature	Units	Description
hometime	minutes	Time at home. Time spent at home in minutes. Home is the most visited significant location between 8 pm and 8 am, including any pauses within a 200-meter radius.
disttravelled	meters	Total distance traveled over a day (flights).
rog	meters	The Radius of Gyration (rog) is a measure in meters of the area covered by a person over a day. A centroid is calculated for all the places (pauses) visited during a day, and a weighted distance between all the places and that centroid is computed. The weights are proportional to the time spent in each place.
maxdiam	meters	The maximum diameter is the largest distance between any two pauses.
maxhomedist	meters	The maximum distance from home in meters.
siglocsvisted	locations	The number of significant locations visited during the day. Significant locations are computed using k-means clustering over pauses found in the whole monitoring period. The number of clusters is found iterating k from 1 to 200 stopping until the centroids of two significant locations are within 400 meters of one another.
avgflightlen	meters	Mean length of all flights.
stdflightlen	meters	Standard deviation of the length of all flights.
avgflightdur	seconds	Mean duration of all flights.
stdflightdur	seconds	The standard deviation of the duration of all flights.
probpause	-	The fraction of a day spent in a pause (as opposed to a flight)
siglocentropy	nats	Shannon's entropy measurement is based on the proportion of time spent at each significant location visited during a day.
circdnrtn	-	A continuous metric quantifying a person's circadian routine that can take any value between 0 and 1, where 0 represents a daily routine completely different from any other sensed days and 1 a routine the same as every other sensed day.
wkenddayrtn	-	Same as circdnrtn but computed separately for weekends and weekdays.

Accelerometer

AWARE Raw Data

RAPIDS column	Stream column
TIMESTAMP	timestamp
DEVICE_ID	device_id
DOUBLE_VALUES_0	double_values_0
DOUBLE_VALUES_1	double_values_1
DOUBLE_VALUES_2	double_values_2

Processed Data

Feature	Units	Description
maxmagnitude	m/s ²	The maximum magnitude of acceleration ($\ acceleration\ = \sqrt{x^2 + y^2 + z^2}$).
minmagnitude	m/s ²	The minimum magnitude of acceleration.
avgmagnitude	m/s ²	The average magnitude of acceleration.
medianmagnitude	m/s ²	The median magnitude of acceleration.
stdmagnitude	m/s ²	The standard deviation of acceleration.

Pandas et al:

sumduration	minutes	Total duration of all exertional or non-exertional activity episodes.
maxduration	minutes	Longest duration of any exertional or non-exertional activity episode.
minduration	minutes	Shortest duration of any exertional or non-exertional activity episode.
avgduration	minutes	Average duration of any exertional or non-exertional activity episode.
medianduration	minutes	Median duration of any exertional or non-exertional activity episode.
stdduration	minutes	Standard deviation of the duration of all exertional or non-exertional activity episodes.

Phone activity recognition (walking, running, cycling, etc.)

AWARE Raw Data

RAPIDS column	Stream column
TIMESTAMP	timestamp
DEVICE_ID	device_id
ACTIVITY_NAME	activity_name
ACTIVITY_TYPE	activity_type
CONFIDENCE	confidence

Processed Data

Feature	Units	Description
count	rows	Number of episodes.
mostcommonactivity	activity type	The most common activity type (e.g. <code>still</code> , <code>on_foot</code> , etc.). If there is a tie, the first one is chosen.
countuniqueactivities	activity type	Number of unique activities.
durationstationary	minutes	The total duration of <code>[ACTIVITY_CLASSES][STATIONARY]</code> episodes of still and tilting activities
durationmobile	minutes	The total duration of <code>[ACTIVITY_CLASSES][MOBILE]</code> episodes of on foot, running, and on bicycle activities
durationvehicle	minutes	The total duration of <code>[ACTIVITY_CLASSES][VEHICLE]</code> episodes of on vehicle activity