2nd Annual Computational Data Neuroscience Symposium
from the Brigham Health/Harvard Dept of Neurosurgery's Computational Neuroscience Outcomes Center & the Harvard School of Public Health Onnela lab
October 23, 2020

Reproducible Analysis Pipeline For Data Streams (RAPIDS): Open Source Software to Process Data Collected with Mobile Devices

Julio Vega

J. Vega; M. Li; K. Aguillera; N. Goel; E. Joshi; K. C. Durica; A. R. Kunta; C. A. Low
University of Pittsburgh

Background

- Smartphone and wearable devices are widely used in behavioral and clinical research (e.g. what is the relation between time at home and depression?)
- Researchers program ad-hoc scripts even though we all use similar mobile sensors, platforms and devices
- The quality of code and software engineering practices vary
- This makes it difficult for other scientists to read, reproduce, reuse, and audit a publication's code and its results

Methods

- RAPIDS is a pipeline to standardize the processing and reporting of data streams coming from mobile sensors.
- Python and R Scripts orchestrated by Snakemake
- Each step in the pipeline is reusable and auditable
- Open source, documented and reproducible
- Its development has been and will be informed by public discussions with the mobile sensing research community

https://www.rapids.science/latest/

Results

- Smartphone data collected with the AWARE Framework on Android and iOS (15 sensors and ~160 behavioral features)
- Fitbit data (sleep, steps, heartrate)
- Parallel execution
- Monitored and unmonitored periods and visualizations for quality control
- Unit tests
- Dynamic day epochs of any length that start at specific times or around certain events
- Single or multiple time zones
- Easy to add new features (R or Python) keeping authorship and citations

Conclusions

- We will add more features, sensors, data cleaning, and analysis scripts
- RAPIDS has reduced the time it took to extract behavioral features from weeks to minutes.
- RAPIDS pipelines can be released along papers (strongly suggested!)
- Open discussions on algorithms on GitHub and Slack

Acknowledgements

Mobile Sensing + Health Institute (MoSHI)

- Meng Li
- Kwesi Aguillera
- Nikunj Goel
- Echhit Joshi
- Krina C Durica
- Abhineeth Reddy Kunta
- Carissa A Low

Julio Vega
vegaju@upmc.edu
@julio_ui
juliovega.info

RAPIDS

https://rapids.science/

Preprint

bit.ly/preprint_rapids

Join the community

#rapids channel in

bit.ly/aware_slack