Project Proposal

5/25/2018

Team: Zachary Stence and Torry Johnson
Mentor: Dr. Bin Wang
Project Description

- **Title:** Determining Whether Two Devices are on the Same Person using Accelerometers
- **Objective:** In our society, wearable devices are becoming more prevalent and useful. Whether it be a smartwatch, fitness tracker, or health device, wearable IoT devices often have no authentication techniques to verify who the user is. This project aims to use machine learning classification algorithms included with the python scikit-learn [5] package to analyze data from wearable device accelerometers. The analysis will conclude with a statistically based binary classification regarding whether or not readings from two accelerometers came from the same person using the device [4]. This could be easily continued into forming an authentication system using accelerometer data.
- **Potential Challenges**
    1. Learning how to effectively use scikit-learn, and learning what functionality included in scikit-learn we need to use
    2. Using the .mat files from the UniMiB SHAR dataset [3] in scikit-learn
    3. Manipulating the data provided into a useful format (extracting the data we need, splitting into training and testing data, learning how to use our own data with the machine learning algorithms in scikit-learn)
    4. Understanding the complicated math behind feature matrices, coherence, and other signal processing techniques
    5. Collecting new data from accelerometers to use in our analysis
- **Potential Solutions**
    1. Follow scikit-learn tutorials, read documentation, and look for tutorials/projects done by other people
    2. Either find a way to convert the data or find another python package that can read matlab data into a useful data structure (scipy.io)
    3. Conceptually figure out what format our data needs to be in, then figure out how to get the data we have to look like that; will likely need to make use of NumPy and/or Pandas
    4. Research signal processing, starting with the fundamentals and building up to what we need to understand
    5. Use apps that log accelerometer values into a spreadsheet
- **Project Plan** [see tables below]

Individual Meetings			
Scheduled Meeting Time	**Attended?**	**Problems If Missed**	**Mentor Agreement/Comments**

May 21st 9:00am	Yes	Reschedule meeting as soon as possible	
May 29th 9:00am		Reschedule meeting as soon as possible	
June 4th 9:00am		Reschedule meeting as soon as possible	
June 11th 9:00am		Reschedule meeting as soon as possible	
June 18th 9:00am		Reschedule meeting as soon as possible	
June 25th 9:00am		Reschedule meeting as soon as possible	
July 2nd 9:00am		Reschedule meeting as soon as possible	
July 9th 9:00am		Reschedule meeting as soon as possible	
July 16th 9:00am		Reschedule meeting as soon as possible	

Group Meetings			
Scheduled Meeting Time	**Attended?**	**Problems If Missed**	**Mentor Agreement/Comments**
May 23rd 10:00am		Reschedule meeting as soon as possible	
May 30th 10:00am		Reschedule meeting as soon as possible	
June 6th 10:00am		Reschedule meeting as soon as possible	
June 13th 10:00am		Reschedule meeting as soon as possible	
June 20th 10:00am		Reschedule meeting as soon as possible	
June 17th 10:00am		Reschedule meeting as soon as possible	
July 4th 10:00am		Reschedule meeting as soon as possible	
July 11th 10:00am		Reschedule meeting as soon as possible	
July 18th 10:00am		Reschedule meeting as soon as possible	

Tasks (gray cells are additional goals to be achieved if time permits)				
Tasks	**Deliverable**	**Completion Date**	**Problems If**	**Mentor**

		Expected	Actual	Missed	Agreement/ Comments
In-depth machine learning research and familiarization	Good understanding of machine learning goals/techniques	5/21	5/21	Trouble understanding breadth/depth of project	
Decide how UniMiB -SHAR dataset will be used or if we will collect and use our own data	Know which dataset(s) we will use and why	5/23	5/21	Need to decide ASAP	Use UniMiB-SHAR
Brush up on Python using online tutorials and coding our own examples	Progress in SoloLearn tutorials	5/23		Working with data may be difficult	
Figure out how to import raw data for use in Python (full_data.mat)	Example program importing a small accelerometer dataset	5/23	5/21	Delay processing, finish ASAP	
Determine features of dataset and how data will be used with machine learning algorithms	Outline of data and how it will be used (including specified features)	5/24		Proposal presentation may lack important info, need to finish ASAP	
Project Proposal Submission	Completed Project Proposal	5/25		Finish ASAP	
Determine what the data should look like ● how will training/testing take place? ● classify data as on same body/not)	Conceptual idea of format of data and how it will interact with scikit-learn	5/28		Proposal presentation may lack important info, need to finish ASAP	
Proposal Presentation	Completed Presentation	5/30		Complete ASAP	
Visualize dataset and features (line plot, boxplots, etc) to gain insight into how	Graphs/tables of information/ statistics about data	6/4		Difficulty determining features of data	

classification will take place					
Get data into correct format (features) ready for use with scikit-learn	Dataset in python in the correct format for scikit-learn	6/8		Finish before train/test deadline	
Outline what our Python code will accomplish	Generic outline of code	6/11		Difficulty coding program	
Code preliminary program in Python for testing one machine learning algorithm	Complete python program	6/14		First training/testing will be delayed	
Train and test our data with at least one algorithm	Results of training/testing (confusion matrix)	6/15		Finish ASAP. Needed by midterm presentation	
Midterm Presentation	Completed presentation	6/20		Finish ASAP	
Midterm Progress Report	Submitted midterm progress report	6/22		Finish ASAP	
Midterm Report	Completed midterm report	6/22		Finish ASAP	
Determine which classification algorithm will work the best with our dataset	List of classification algorithms ranked by train-time, test-time, and score (confusion matrix)	6/25		Finish soon, may run out of time	
Tweaking dataset/algorithms used to solve any unexpected issues that may arise	Fully functional algorithm operating on data	7/7		Finish ASAP to allow for more in depth analysis	
Find best classification algorithm(s)	Confidence level of authentication based on length of data	7/9		Finish ASAP	

Obtain statistics about data and algorithm (visualizations) ● Confusion matrix ● How much time is needed for authentication? ● How reliable is authentication? ● How do different algorithms perform?	Graphs/plots and information/ statistics about data, algorithms, and overall project	7/11		Finish ASAP for final report	
Collect and use our own data	Full dataset of our data collected	N/A		Finish if time permits	
Further optimize analysis by refining data and/or algorithms used (hyperparameters)	Machine learning model with improved functionality	N/A		Finish if time permits	
Obtain statistics after optimization and observe the increase in performance	Graphs/plots and information/ statistics about data, algorithms, and overall project	N/A		Finish if time permits	
Create an example authentication app based off of data collected and analyzed in real-time [7]	Working app prototype	N/A		Finish if time permits	
Final Report/Presentation	Completed final report/presentation	7/18		Finish ASAP	
Poster Presentation	Completed poster	7/19		Finish ASAP	

Resources Needed (hardware, software, data, other)				
Resources Needed	**Available Time**		**Problems if delayed?**	**Mentor Agreement/ Comments**
	Expected	**Actual**		
UniMiB SHAR dataset	Now	Now	Already have data	
Smartphone	Now	Now	Finish if time permits	

Smartwatch	Now	Now	Finish if time permits	
Sensor Kinetics iPhone App	Now	Now	Finish if time permits	
Power Sense iPhone App	Now	Now	Finish if time permits	
Android App	Now	Now	Finish if time permits	
Manually collected dataset	N/A	N/A	Finish if time permits	

References

[1] Andrea Bianchi, Ian Oakley. "Wearable authentication: Trends and opportunities" Usable privacy and security / Alexander De Luca, Emanuel von Zezschwitz. it - Information Technology, 58.5 (2016): 255-262. Retrieved 25 May. 2018, from doi:10.1515/itit-2016-0010

[2] Cory T Cornelius, David F Kotz. “Recognizing Whether Sensors Are on the Same Body.” Pervasive and Mobile Computing, vol. 8, no. 6, Dec. 2012, pp. 822–836. ScienceDirect, doi:10.1016/j.pmcj.2012.06.005.

[3] Daniela Micucci, Marco Mobilio, Paolo Napoletano. “UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones.” Applied Sciences, vol. 7, no. 10, 24 Oct. 2017. Applied Sciences, doi:10.3390/app7101101.

[4] Prettenhofer, Peter. “Classification with Scikit-Learn.” DataRobot Blog, 3 Mar. 2014, blog.datarobot.com/classification-with-scikit-learn.

[5] “Scikit-Learn: Machine Learning in Python.” Scikit-Learn, Oct. 2017, scikit-learn.org/stable/index.html.

[6] Taspinar, Ahmet. “Classification with Scikit-Learn.” Ahmet Taspinar, 1 Mar. 2018, ataspinar.com/2017/05/26/classification-with-scikit-learn/.

[7] Xu, Weitao. “Gait-based authentication system on smart wearable devices.” Online video clip. YouTube. YouTube, 22 Aug. 2016. Web. 21 May 2018.