Gerontechnology I

Fall 2017

Homework Assignment

Due date: ***Thursday, October 12, 2017, by 9am (Email deadline)***

Description

The purpose of this homework assignment is to give you hands-on experience with collecting activity data and analyzing the data. You will have an opportunity to work with mobile apps, apple watches, sensor data, and a machine learning toolkit.

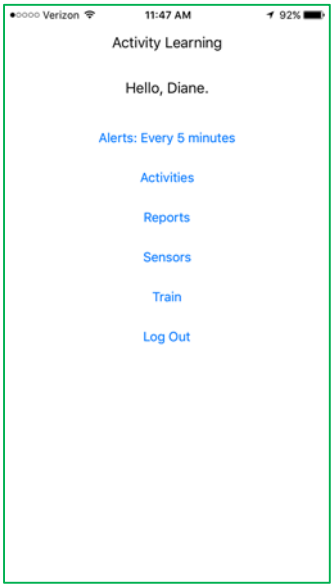
There will be three phases of this project: 1) data collection, 2) data analysis, and 3) report generation.

1. Data Collection

You will be given a smart phone or 2 smart watches to keep with you for two weeks. Alternatively, if you currently own a recent-version iPhone or Android phone, you can load the app onto your existing phone using the instructions found online at http://eecs.wsu.edu/~cook/gt1/hw. If you borrow our phones or watches, please be sure to bring it back to class within three weeks. The phone / watch has an app preloaded on it called "Activity Learning" (AL). It looks like this:

This app tries to learn activities that you regularly perform. To do this, it will periodically query you (via a "notification"). For five seconds before the query, it collects and stores data using sensors available on the phone. The collected sensor data is sent periodically to a machine on campus so that the activities can be learned. Eventually it may become smart enough to recognize the activity you are performing based on the nature of that sensor data. However, all we want to use AL for is to collect sensor data and activity labels that ***you*** can use to analyze.

The first time you start up the app you will need to do some setup. You will need to tell the phone it is "OK" to allow notifications. You will also need to click on "New User" and provide a name for yourself (your actual name or a made-up name), an email address, and a password. As soon as you do this, you will receive an email message at the address you provided. Follow the instructions provided in that email address to complete registration. After that, go back into the app and "Log in" using the name and password that you provided. Finally, send email to reanne.cunningham@wsu.edu with the name and email address you used, so that we can send you the sensor data you collect.



Now that you are all set up on the app, you can use it to collect sensor data and labels for the activities you perform. We would like you to collect data for five activities: ***Eat, Hobby, Walk, School, and Sleep***. On the main page of the app (shown below on the left) select "Train" and then select which of the five activities you will perform. You will need to then perform that activity for at least the next five seconds.

In order to collect a sufficient amount of data for the machine learning algorithms, you need to collect twenty data points for each of the five activities, for a total of at least one hundred data points. In order to get the most of this assignment and the analysis, we encourage you to collect data points on different days throughout the week but to perform the activities using the time, place, and method you would normally employ to perform each activity. This data collection step must be completed no later than class time on ***Tuesday, September 19, 2017***. This is important so that 1) we get the devices back for use in our studies and 2) we can format and send you your collected data for analysis. If you want to see how many data points you have completed for each activity so far, go to the app's main page, select Reports, and look at the Activity Bar Chart.

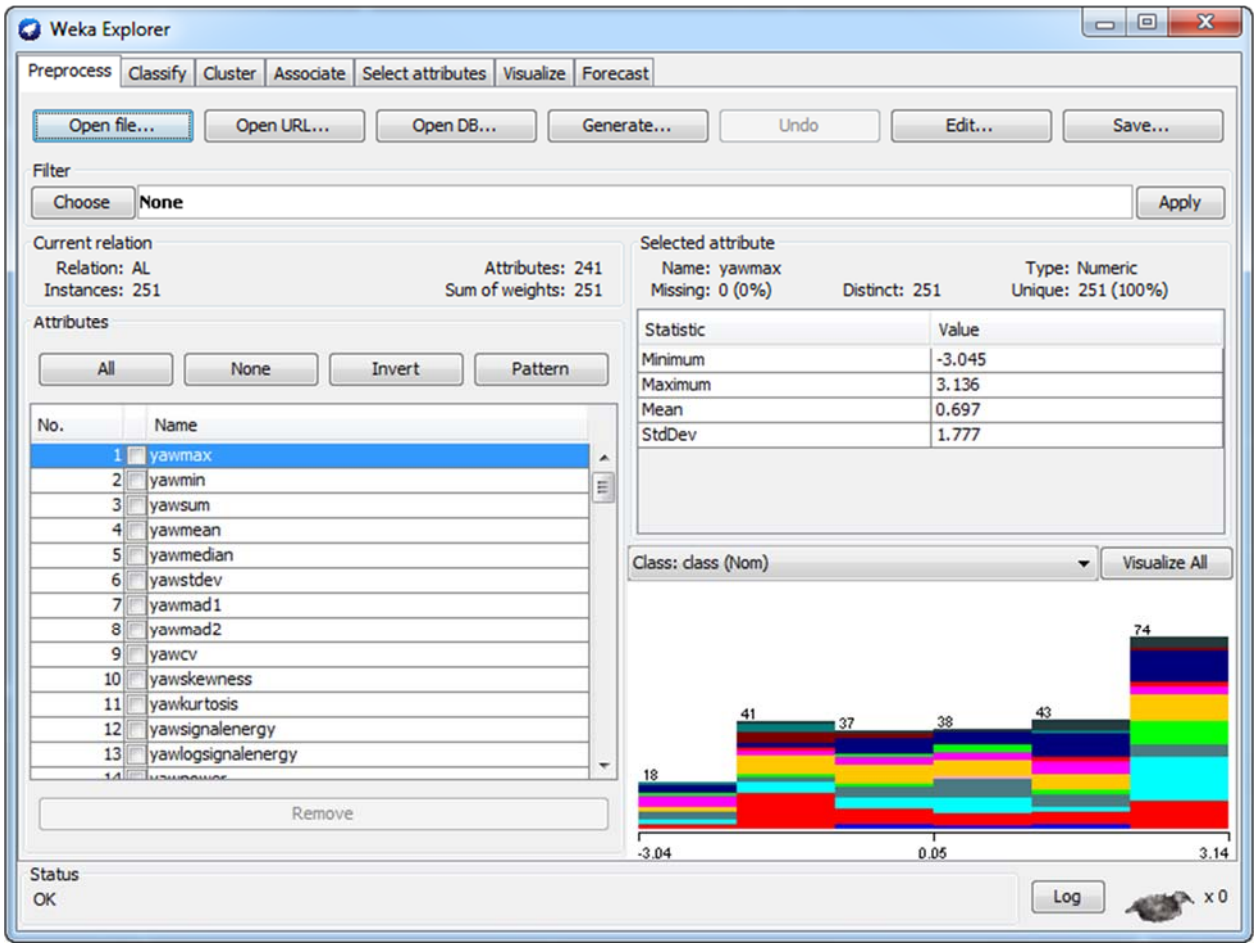
On the main AL page, if you click on "Sensors", you can see types of sensor data that is collected. Try moving the device around and see what happens to the values for acceleration and rotation. Also try walking around and see what happens to the values of latitude and longitude.

2. Data Analysis

We will mail you (using the email address you provide) two files, called data1.arff and data2.arff. The files will be mailed to you no later than Thursday, September 21, 2017. We will not mail you the raw sensor data. Instead, we will be extracting features of the sensors and provide these values to you. We will discuss these in class, but they include max, min, sum, mean, median, standard deviation (std), mean absolute deviation (mad1 and mad2), coefficient of variation (cv), skewness, kurtosis, signal energy, log signal energy, power, and autocorrelation, applied to the different sensor types (yaw, pitch, roll, rotation x (rotx), rotation y (roty), rotation z (rotz), acceleration x (accx), acceleration y (accy), acceleration z (accz), location type (ltype), course, speed, and the type of device that was used to collect the data (dtype). The first file, data1.arff, contains data specifically for you. The second file, data2.arff, contains data for the entire class except for you.

This file is in a format that can be processed by the machine learning toolkit called Weka. You will need to download the Weka software onto a machine. There are versions for Linux, Windows, and Mac machines available from the Weka web page at http://www.cs.waikato.ac.nz/ml/weka/ (click on Download). We will provide a tutorial in class of how to use Weka and there is a tutorial available on the Weka web page. When you start up Weka you will see this window:

Click on Explorer to get the main interface window. On the Preprocess tab, click on Open File and select your data1.arff or data2.arff file. You will now see a window like the one below. All of your features are listed on the left, with your "class" feature (the activity) listed at the bottom. See what happens when you click on different feature names. If you click on the class feature, you will see how many data points you have for each class. See also how the distribution of feature values is different for each (color coded) activity class.



The goal of this step is to try different machine learning algorithms on this dataset and report their performance, making as many observations as you can about the nature of the data, the types of learning algorithms that perform best, and situations that create challenges for the learning algorithms. Weka has many machine learning algorithms in its software set that should be able to learn a mapping from the sensor features to the activity label. To try them out, click on the Classify tab at the top of the screen. You can choose a classifier on this screen. You can also specify how you want to test the classifier (cross validation, percentage split, or other methods). Once you make your selection, click "Start" and Weka will train and test your

classifier on this data. A number of performance measures will be reported including accuracy, true positive rate (TP), false positive rate (FP), and ROC area. It will also report a confusion matrix, which shows for each activity class how those data points were labeled.

3. Report

To complete the homework assignment, create a report detailing your observations of the data collection process and the data analysis process (with results). The report should be a minimum of three pages but should not need to be longer than ten pages. The report should include the following information:

- Observations of how easy / difficult AL was to use
- A discussion of how easy / difficult AL might be for older adults to use
- Using 3-fold cross validation, report classification performance on your own data (data1.arff). For this, pick at least one rule-based classifier, one function-based classifier, and one tree-based classifier.
    - For the tree-based classifier, look at the model that was built. What is one of the rules that was learned? Does it make sense?
    - Report performance in terms of accuracy and area under the ROC curve. Are these results surprising? Are they better than random guessing?
    - Summarize the confusion matrix. What activities are easy to detect and which created difficulty?
- Using the classifier method that performed the best in the previous step, now train a model using data from the rest of the class (data2.arff) and test it on your own data (data.1arff).
    - Report performance in terms of accuracy and area under the ROC curve. How do these results compare with cross validation on your own data?
    - What do these results tell you about the ability to generalize a model for multiple individuals?
- Conclude with final observations and suggestions for improving the app and/or the classifier. Discuss how this could be used for health-assistive technologies.

Turning in Your Homework

If you have questions, please direct them to Reanne Cunningham (reanne.cunningham@wsu.edu). The final homework should be emailed as a pdf file no later than 9am on ***Thursday, October 12, 2017***.

Important Dates

- August 29: Obtain phone / watch with AL app
- September 19: Deadline to finish data collection
- September 21: Receive data to analyze
- October 12: Homework due