The Identification of a Subject by the Acceleration of His Walking

Eyal Gon, Golan Kai & Kotzer Dan
Supervised by Shlomo Rozenfeld

Introduction

Humans have a unique pattern of walking; this pattern tags a person like a fingerprint. Therefore it is possible to identify the subject by his walking. A human's eye would barely be able to notice the minute differences between two people walking, however, operating with an artificial intelligence makes it much easier.

In our experiment we developed a software program that examines his walking and then classifies the subject. Afterwards we examined two learning methods:

- Logistic Regression
- Neural Network

Research Method

We gathered subject's three-dimensional acceleration by attaching a phone to their back; we used this phone as a motion sensor:

We used Python with these packages:

- Numpy for calculating
- Matplotlib for graphics
- Sklearn for the Logistic regression
- Pybrain for the Neural Network and for defining the database.

We separated each sampling of a minute into parts of 2 seconds each, in order to have more examples for the machines learning process. From each part we extracted several features- maximum, average accelerations and histogram for each axis.

After the database was processed and ready to be used it was separated into 3 groups:

- Training for the machine
- Validation to measure the error of the machine
- Test group for the best network that we found

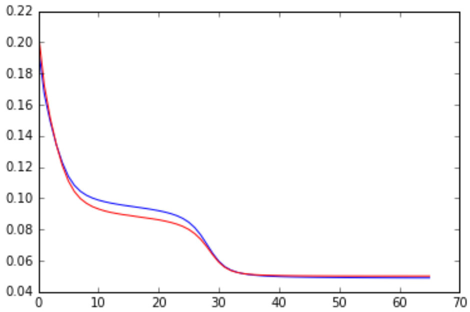
We wrote our code as "Auxiliary Functions" code instead of one long complex solidity, which is essential for understanding and debugging.

For each person we created a Logistic Regression, which identifies his features. Then they tried to define their subjects from the validation group and each printed its error. We combined the errors and got the final result.

Moreover we created an artificial Neural Network, which acts as the neural network in the brain, with one hidden layer. The amount of entrances in the input layer equals the number of features that we decided to use in each experiment. The size of the output layer was determined as the number of subjects that was to be classified that time. We used one hidden layer, which analyzes the connection between the input and the output layer, the amount of neurons was changed in a certain range to find the best net using "For" loop.

Results

There were tests which had 0 % error but we can't assume that the program is 100% right because the size of the database was comparatively small. After a huge amount of runs we found out the best Neural Network has 44 neurons in the hidden layer, the result was 85% right. The Logistic Regressions best result, on 3 subjects, was 96.5% success. Here is an example of the "percentage of error per period of learning", the red represent the training group and the blue the validation group.



The next step was to see the influence of some features on the learning process. The Neural Network without the histogram feature hit 0% error many times, with 26 neurons in the middle layer.

We decided to continue researching with the Neural Network method it was comfortable and easy to modify. This time we expanded from 3 to 9 subjects, without the histogram feature. The best net had 35 neurons in the hidden layer and identified right 42% of the time.

Conclusions

It is important to remember that a guess between 3 people would be 33% success rate and our net has had 85% success rate. In addition a guess of 9 people would be 11% right and we had almost 4 times more accuracy.

By looking at our results we can be sure that our program has a significant learning ability. We think that if we were to us a bigger database we would hit better success rates.

Future Uses

We believe that our program has a lot of potential in other fields: identifying a walking pattern that damages one's body or a thief with a stolen phone and even someone under alcohol/drugs. All of the above could be achieved by a simple phone application based on our work.