Typeguess: Using Mobile Typing Dynamics to Predict Age, Gender and Number of Fingers Used for Typing

Baris Akis, Mariano Sorgente, Abraham Starosta

December 12, 2014

Abstract

We aim to predict individual's age, gender and number of fingers they used to type based on their typing on mobile phones. Although similar work has been done predict a user's information, it has been based on either content of the text or typing style on keyboards instead of analyzing the user's movement patterns. We created our own dataset by building an iOS application we called Typeguess, which allowed us to collect data in person. We then generated 29 features to observe. We did feature analysis for each classifier and followed it with model selection. Finally we were able to select top features and the best working models for classifying age, gender and number of fingers used to type. Best results achieved were respectively 0.752, 0.603, 0.809 on balanced accuracy.

1 Background

Personal identifiers are a subset of of information that alone or combined with certain information can identify a unique individual. Most of the current PIDs are either biometric such as fingerprint or based on assigned numbers such as SSN. A promising area of research is discovering human behavior that can be used for personal identification such as stepping behavior. The use of technology is starting be tested as a personal identifier as well. The field looking into computer keyboard usage is keystroke dynamics.

Keystroke Dynamics use manner and rhythm which individuals types on keyboards as personal identifier. Most of this research was based on static keyboards and used keyboard strokes, duration and latency as features.

Keystroke Dynamics based Authentication (KDA) related work was also conducted for mobile phones. Most of this research was based on decreasing errors for user authentication and they were conducted on handheld phones which didn't provide sensor information.

2 Introduction

Different from the previous work on KDA our project was focused on predicting actual information about the users based on the keystroke dynamics and phone sensor information. With our project TypeGuess we aim to predict individuals typing style (number of fingers), age and gender as they type.

3 Dataset

3.1 Data collection

We created a mobile application to collect the data. The application presents three paragraphs and we ask the user to type the text on the screen as fast and as accurate as possible into the input box. Over the term of this quarter we have collected data in person using our mobile data collector application. The data collection is done in person using iPhone6s as our data collection device. We decided that we should keep the type of collection device constant and we shouldn't use a device that certain users would be over adapted to. So we decide to use iPhone6. We stored keystroke data with additional accelerometer and gyro vectors. We also gathered subject's following personal information: phone size, use of autocorrect, age, gender, mother tongue, and number of thumbs their use to type.

Over the time of 2 weeks we collected 132 samples with the following breakdown:

Property	Frequency
Male	0.52
Female	0.48
<30 years old	0.68
>30 years old	0.32
1 finger	0.184
2 fingers	0.816

3.2 Data Processing

After collecting a data point, it is sent to our python server where we store it for analysis. The data was stored in JSON format, and passed in to a python program which extracts the features to a feature matrix for each subject.

3.3 Generating features

We collected age, gender, number of fingers used, regular phone used, size of phone, autocorrect, and mother tongue as possible predictable values. Later we decided the concentrate only on age, gender and number of fingers used.

Thinking about different possible features we can extract from key logs and some online research based on other studies we came up with 29 different features to used. These features include:

- Speed based features such as total completion time, average time between keys
- Accuracy based features such as total keys pressed, number of backspaces used, edit distance to original text
- Special keys based features such as the time to find shift key and insert punctuation
- Accelerometer information on x, y, z axises such as change in accelerometer when hitting "next" and also average accelerometer value through typing
- Gyrometer information on x, y, z axises such as change in accelerometer when hitting "next" and also average accelerometer value through typing

4 Feature Selection

4.1 Defining accuracy

$$BalancedAccuracy = \frac{.5 * TP}{TP + FN} + \frac{.5 * TN}{TN + FP}$$

Where TP = number of true positives (eg. predicted old, actual old)
TN = number of true negatives
FP = number of false positives
FN = number of false negatives

We used balanced accuracy as our standard to choose the best features. This is because since our data was skewed (more young than old, more two fingers than one finger), simply computing the total accuracy gave us misleading results.

4.2 Features overview and Selection

We currently have 29 different features. These features can be categorized under 5 main different groups: Time, content accuracy, key presses, special keys, accelerometer and gyroscope. We selected top 5 features for each of our classifiers and used those for model selection. We tried to run PCA but it didn't really help to our app besides better visuals for graphing. Selecting best features we ran each of them one by one with all possible models. Below you can find top features for each classifier with the balanced accuracy they returned when the models were run only with the selected feature. We ran all features alone using 0.1 left out CV with 7 different classifiers. We selected the top 5 for each.

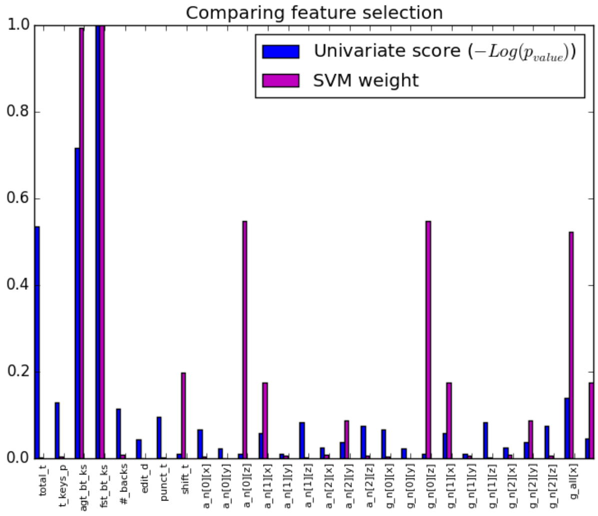
4.3 Number of fingers

Best features (Balanced Accuracy):

- Fastest time between keys, 0.779
- Average time between keys, 0.695
- Total time, 0.649
- Gyroscope Y coordinates, 0.634
- Gyroscope Z coordinates, 0.623

Statistics:
Top Model:Decision Tree

Figure 1: Fingers feature selection



Balanced Accuracy: 0.809
Testing error: 0.191
Training error: 0.134

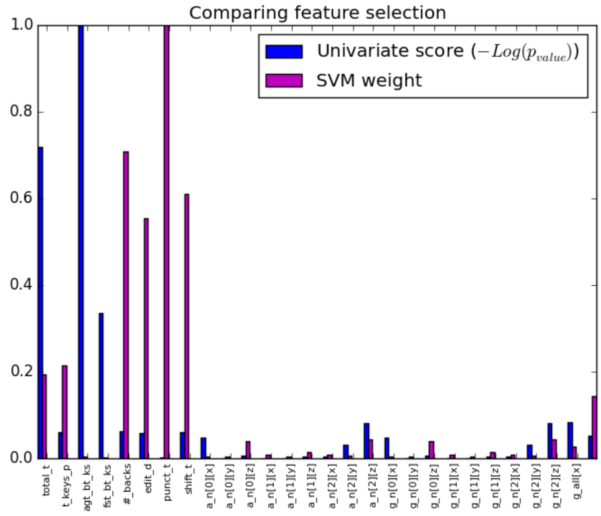
Number of thumbs turned out to be the easiest value to predict correctly. It is also highly correlated with age (older people tend to use 1 finger more often than younger people). This also explains why average time is up there as a top feature. The most useful feature for this prediction turned out to be fastest time between keys. This is the average of the top 10 fastest key press combinations. It is likely that when typing with one finger, it is difficult to type two keys simultaneously at a fast speed, while it is easy and more common while typing with two fingers. Also we found that the gyroscope data was different for one finger typists. Our suspicion is that holding the phone with one hand leads to different holding angles, and thus these features were useful.

4.4 Age

Best features (Balanced Accuracy):

- Average time between keys, 0.709
- Fastest time between keys, 0.678
- Total time, 0.673

Figure 2: Age feature selection



- Time to press 'shift' key, 0.611
- Number of backspaces, 0.578

Statistics:
Top Model: SVC Linear Kernel
Balabced Accuracy: 0.752
Testing error: 0.248
Training error: 0.126

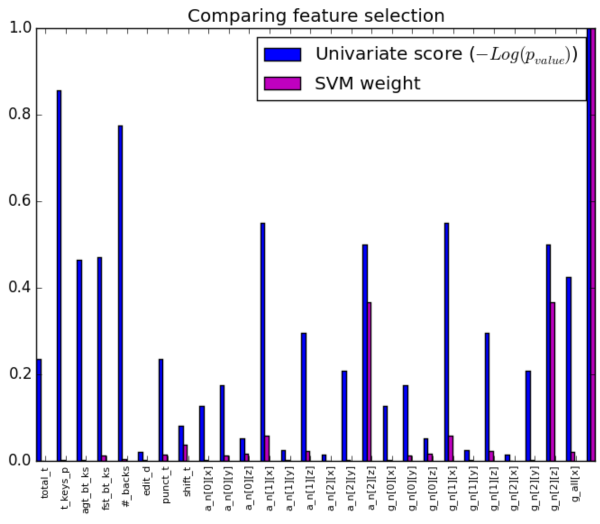
Finally, predicting age was our top priority for this project. We expected that older people would take a significantly longer time to type in the text. Our results confirmed this, as the top four features are related to typing speed. With just average time between keys, and shift time (time taken to press shift), we were able to capture most of the observable differences between the age classes, and achieve acceptable high accuracy scores.

We decided to classify ages between two groups: older and younger than 30. In order to pick this threshold we took into account both the accuracies of each model with different age thresholds. Furthermore, we also tried to balance it by trying to keep the groups (older and younger than 30) with similar number of training examples.

4.5 Gender

Best features:

Figure 3: Gender feature selection



- Edit distance(User Input, Shown Text), 0.600
- Gyrometer X coordinates, 0.583
- Number of backspaces, 0.581
- Total keys pressed, 0.576
- Gyrometer Y coordinates, 0.572

Statistics:
Top Model: SVC Linear Kernel
Accuracy: 0.603
Testing error: 0.397
Training error: 0.303

For gender Logistic Regression came out as the most predictive model. Although 0.603 is not an extremely high number, we were pleased to be able to distinguish males and females by even a small margin. It turns out that the most useful feature here was edit distance (The number of modifications to the person's input required for it to be equal to the prompt text). It turns out that males tend to make more mistakes while typing, or they simply make less backspaces to correct input. Another useful feature was the x direction gyroscope reading. According to our findings, females tend to hold the phone at a slightly higher angle; it could be related to the person's height, or maybe the way in which the male pose tends to differ from the female's.

4.6 Model selection

For selecting optimal models, we ran each model in the following list with our top features for each classifier. Interestingly enough, our three different Y variables were each predicted best with a different model.

These are the ones we tested:

- KMeans with 2 clusters
- Quadratic Discriminant Analysis
- Linear Discriminant Analysis
- Decision Tree Classifier, depth of 5
- K Nearest Neighbors Classifier
- Support Vector Machine Classifiers (with several kernels)
- Gaussian Naive Bayes
- Logistic Regression

In the next table you can see the scores we got from each model. These scores were obtained by running the models with the best parameters we could find, and using the top five features for each Y. As you can see, each prediction was optimized with a different model:

Model	Age score	Gender score	Finger score
SVC (linear)	.752	.573	.592
Logistic R.	.751	.603	.596
Gaussian NB	.741	.556	.727
K Neighb.	.712	.483	.635
Decision T.	.711	.550	.809
QDA	.698	.546	.749
LDA	.684	.572	.700
SVC (quadr)	.519	.517	.564

4.7 Unsupervised learning

Although all our data was labelled, we thought we might attempt to run some unsupervised learning algorithms on our dataset. Specifically, we ran K-means clustering and Affinity Propagation. K-means clustering with 2 clusters resulted in satisfactory results: around 0.69 accuracy for predicting age and 0.72 accuracy for predicting number of fingers. Although we could get some use from these results, they were inferior to most of the supervised algorithms that we ran. Affinity propagation, like several other clustering algorithms, is meant for finding many more than two clusters, so it was not fit for our two bucket classification problem. The results we managed to get here were unsatisfactory, so we stuck with our best supervised models.

4.8 PCA

After selecting most of our features and just before implementing model selection we wanted to check the effect of PCA so we also ran many models with PCA as well. Though the difference in accuracy was negligible. For some models PCA had a small positive effect, while for others it had a small negative effect. As we talked in class PCA is much suitable for visualization or optimization in large datasets. We also attempted using AdaBoost, but this didn't really improve our performace either.

5 Conclusion and Future work

Overall, our work indicates that it is possible to predict a person's information based on mobile typing dynamics. We believe that our current work can serve as a stepping stone.

We believe that keystroke dynamics combined with sensor information has potential to predict significant amount of personal information. In order to move this research forward there is a huge need for more data. The data needs to be more diverse and more detailed. Currently we have typing info of less than three minutes. This can be significantly expanded by logging keystrokes during daily uses. Additionally it's possible to extract even more features using latent feature models. A different approach to this problem might be using deep learning if high amounts of user data is acquired.

6 References

- Jaeseok Yun et al, Int. J. Patt. Recogn. Artif. Intell. 22, 497 (2008). DOI: 10.1142/S0218001408006338
- Deng Y., Yu Y., Keystroke Dynamics User Authentication Based on Gaussian Mixture Model and Deep Belief Nets, ISRN Signal Processing Volume 2013 (2013), Article ID 565183, http://dx.doi.org/10.1155/2013/565183
- N. L. Clarke, S. M. Furnell. Authenticating mobile phone users using keystroke analysis, International Journal of Information Security January 2007, Volume 6, Issue 1, pp 1-14
- Improving Authentication Accuracy Using Artificial Rhythms and Cues for Keystroke Dynamics-based Authentication. N.p., n.d. Web. 12 Dec. 2014.
- Advanced User Authentication for Mobile Devices. N.p., n.d. Web. 12 Dec. 2014.
- Keystroke Dynamics in Password Authentication Enhancement. N.p., n.d. Web. 12 Dec. 2014.