Keystroke Dynamics

CS725 Project

Indian Institute of Technology, Bombay
Department of Computer Science and Engineering

Devashish Singh (163059001)
Prateek Patidar (163059006)
Shubham Singh (163059008)
Hareesh Kumar (16305R013)

1 Project Description

Keystroke dynamics is the study of whether people can be distinguished by their typing rhythms, much like handwriting is used to identify the author of a written text. Possible applications include acting as an electronic fingerprint, or in an access-control mechanism. A digital fingerprint would tie a person to a computer-based crime in the same manner that a physical fingerprint ties a person to the scene of a physical crime. Access control could incorporate keystroke dynamics both by requiring a legitimate user to type a password with the correct rhythm, and by continually authenticating that user while they type on the keyboard.

2 Tentative Approach

We'll proceed using the following workflow:

- (training): Retrieve the first 200 passwords typed by the genuine user from the password-timing table. Use the anomaly detector's training function[1] and other functions with these password-typing times to build a detection model for the user's typing.
- (cross validation): Retrieve the last 200 passwords typed by the genuine user from the password-timing table. Use the anomaly detector's[1] scoring function and the detection model (from Step 1) to generate anomaly scores for these password-typing times. Record these anomaly scores as user scores.

  Repeat the above four steps, designating each of the subjects as the genuine user in turn, and calculating the equal-error rate for the genuine user. Calculate the mean of all 51 subjects' equal-error rates as a measure of the detector's performance, and calculate the standard deviation as a measure of its variance across subjects.

3 Papers

- Comparing Anomaly Detectors for Keystroke Dynamics[1]
- ROCR: visualizing classifier performance in R [2]

4 Datasets

We will be using dataset from CMU. The dataset consists The data consist of keystroke-timing information from 51 subjects (typists), each typing a password (.tie5Roanl) a total of 400 times in 8 sessions. The dataset consists of digraph data for keystrokes.

5 Work Done till Now

We did some feature engineering on the dataset and used some classification methods available in scikit-learn library on the data like Logistic Regression, Support Vector Machines, Random Forests, K NN Classification and Gaussian Naive Bayes.

References

[1] Kevin S. Killourhy and Roy A. Maxion. Comparing anomaly detectors for keystroke dynamics. In *Proceedings of the 39th Annual International Conference on Dependable Systems and Networks*, DSN-2009, pages 125–134, New York, NY, USA, 2009. IEEE.

[2] Tobias Sing, Oliver Sander, Niko Beerenwinkel, and Thomas Lengauer. Rocr: visualizing classifier performance in r. *Bioinformatics*, 21(20):3940, 2005.