**Report on Term Project**

**Machine Intelligence and Expert Systems**

**Group - 3**

**Divyansh Jhunjhunwala(16EC10066)**

**Siddharth Das(16EC10067)**

**Pratik Tibrewal(16EC10065)**

**Umang Agarwal(16EC10056)**

**Vaibhav Krishna(16EC10064)**

**Problem Statement-** Authenticate User by analyzing keystroke dynamics data using classifier based on Self Organizing Maps

**Introduction-** In this project, we treat Keystroke dynamics as behavioral biometrics to authenticate a user. We use 2 features to authenticate users, namely, the hold time of individual key and the latency of the consecutive keystrokes.

The data was collected over a period of 5 weeks using the keylogger software and later preprocessed to give .txt files with hold time and latency times between consecutive keystrokes for all key pairs for a user.

**Methodology-** For classifying the data, we used the Self Organizing Map model which is a type of ANN trained by unsupervised learning. We train our model with the data of 8 users who are supposed to be authentic and test the model for 2 users and try to find out if the model can recognize them as intruders. The features are represented in 2 ways, i.e., as (<key pressed>, <hold time>) and as (<successive presses>,<latency>).

The model is then trained using these features for a training size of about 10000 examples and then the errors are quantized. A threshold of 80% is put on the data to recognize outliers and thus identify intruders. We show the result in terms of True Positive(TP), False Positive(FP), True Negative(TN) and False Negative(FN).

**Results and Analysis-**

Training parameters

1. Grid size- 50\*50
2. Distance function-Gaussian
3. No. of iterations-10000

Error Threshold for anomaly- 80th percentile of quantization error for training data

Fold 1:

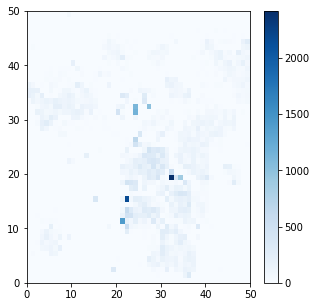
No. of training examples-100000

Test:

Authentic - 11872 examples

Intruder- 34674 examples

Node map



|  |  |  |  |
| --- | --- | --- | --- |
|  | Actual Label | | |
| Predicted Label |  | True User | Intruder |
| True User | 8258 | 27165 |
| Intruder | 3613 | 7509 |

Fold 2:

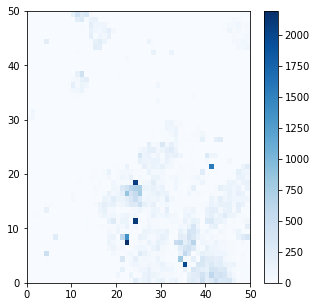
No. of training examples-100000

Test:

Authentic - 6474 examples

Intruder- 40072 examples

Node map



|  |  |  |  |
| --- | --- | --- | --- |
|  | Actual Label | | |
| Predicted Label |  | True User | Intruder |
| True User | 4346 | 16462 |
| Intruder | 2127 | 23619 |

Fold 3:

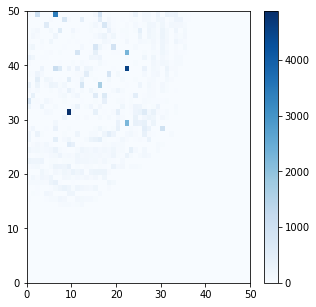
No. of training examples-110000

Test:

Authentic - 16882 examples

Intruder- 19664 examples

Node map



|  |  |  |  |
| --- | --- | --- | --- |
|  | Actual Label | | |
| Predicted Label |  | True User | Intruder |
| True User | 11993 | 14400 |
| Intruder | 4888 | 5264 |

Fold 4:

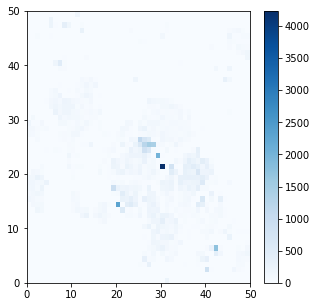
No. of training examples-110000

Test:

Authentic - 17204 examples

Intruder- 19342 examples

Node map



|  |  |  |  |
| --- | --- | --- | --- |
|  | Actual Label | | |
| Predicted Label |  | True User | Intruder |
| True User | 11381 | 12637 |
| Intruder | 5822 | 6705 |

Fold 5:

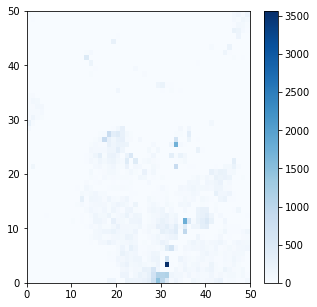
No. of training examples-100000

Test:

Authentic - 13752 examples

Intruder- 32794 examples

Node map



|  |  |  |  |
| --- | --- | --- | --- |
|  | Actual Label | | |
| Predicted Label |  | True User | Intruder |
| True User | 8753 | 21060 |
| Intruder | 4998 | 11734 |

**Conclusion-**

* The high false positive rate of our classifier can be attributed to the feature vector size of 2. It appears that for the SOM classifier the feature vector does not capture sufficient information which can help distinguish it between an authentic user and fake user.
* As an alternative we can consider using the feature vector of length 2*N* − 1, where *N* is the number of keystrokes. *N* keystrokes along with *N* − 1 number of latency generate feature vector of length 2*N* − 1. This increase may be helpful for better discrimination as it captures the chronology of the event. This is suggested in the paper “User Authentication Based on Keystroke Dynamics”.

**References-**

1. Rajat Kumar Das, Sudipta Mukhopadhyay, and Puranjoy Bhattacharya. "User Authentication Based on Keystroke Dynamics." IETE Journal of Research 60.3 (2014): 229-239.

2. Abhinav Ralhan. “Self Organising maps.” Towards Data Science, 2018

3. Rajat Kumar Das, Sudipta Mukhopadhyay and Puranjoy Bhattacharya "Continuous multimodal biometric authentication for PC and handheld devices", IETE Journal of Education, pp 59-69 vol. 52, Issue 2, July-December 2011.