Enhanced Password System

The George Washington University

CSCI 6364 _Machine Learning

Linge Yan

1. Introduce

Identity authentication technology based on human biological characteristics is a hot spot in the field of information security technology. Due to the uniqueness of human biological characteristics, it is safer and more reliable to apply human identification. With the widespread popularization and application of mobile terminals such as smart phones, people are increasingly using it for various transactions that require authentication, such as bank payment, air ticket ordering and payment insurance, so that smartphones and various types of mobile terminals are authenticated Become an important guarantee for secure payments. The traditional way of adding a simple user name and password has the risk of being easily cracked. Therefore, adopting a human-based biological feature is of great significance for improving the security of such payment. Based on human biological characteristics of identity authentication are based on fingerprints, face, iris and other characteristics of the identity authentication. However, when they are applied to mobile terminals such as smart phones, the problems of smart phone resources or the complexity of the identification algorithm itself may lead to problems of low identification efficiency and incorrect operation. This project is intended to realize the biometric feature of extracting more users through mobile phones and to use the classifier to achieve more accurate and efficient security identification technology.

2. Feature selection

We have made Android APP, which is used to collect the user's relevant characteristics when entering the password. By calling Android's existing related API, you and I are very convenient to get a lot of touch character data. Specific features are as follows:

Fig.1 The App to collect the data and feature

1) Acceleration:

For every number d in a pin action, we calculated five acceleration values:

a1: The magnitude of acceleration when the number d is depressed;

a2: The magnitude of acceleration when the number d is loosened;

a3: The maximum value of acceleration when the number d key is released;

a4: The minimum value of acceleration when the number d key is released;

a5: The mean value of acceleration when the number d key is released.

The data obtained through the following operation to get a unique feature value:

$$\|\vec{a}\| = \sqrt{a_x^2 + a_y^2 + a_z^2}.$$

We chose not to use a separate component because the phone coordinate system is sensitive to changes in position. Similarly, calculate the characteristics of angular acceleration in a similar way. In combination with acceleration and acceleration related features, there are a total of 40 in four-digit pin movements.

2) Pressure:

We use Android API motionevent.get to get pressure readings. The returned pressure measurement is an abstract unit from 0 (no pressure) to 1 (normal pressure), but values above 1 may appear depending on the calibration of the input device (according to the Android API documentation). In this feature set, we include pressure readings for the release of keys and keys. A 4-digit password has 8 pressure-related features.

3) Size:

Similar to the pressure reading, another Android API calls motionevent.getsize () to measure the touch size associated with each touch event. According to the Android document, it returns a scaled value of the approximate size of the given pointer index. This represents the approximation of the screen area being pressed. The actual value of the pixel corresponding to the touch is normalized with a specific range of the device and is scaled to a value between 0 and 1. For each button and key release, we record the size reading and are included in the feature set. A 4-digit pin contains 8 size-related features.

4) Time:

Key time and interval between two adjacent keys. They are measured by the timestamp of the touch event, they are both touch and touch. In general, a 4-digit PIN action contains seven time-related features.

Feature	Description	Dimensions(4-digit password)	API
Acceleration (linear & angular)	At TouchDown	8	getDefaultSensor(Sensor.TYPE_ACCELEROMETER)
	At TouchUp	8	
	Min in key-hold	8	
	Max in key-hold	8	
	Mean in key-hold	8	
Pressure	At TouchDown	4	MotionEvent.getpressure()
	At TouchUp	4	
Touched Size	At TouchDown	4	MotionEvent.getsize()
	At TouchUp	4	
Time	Key hold time	4	TouchEvent()
	Inter-key time	3	
Total	All features	63	

Table. 1 The features in this paper

After repeated verification, we determine that these features are extremely stable. As shown in Figure 3, even if the touch screen is in different posture, the same person's corresponding features refer to the same range.

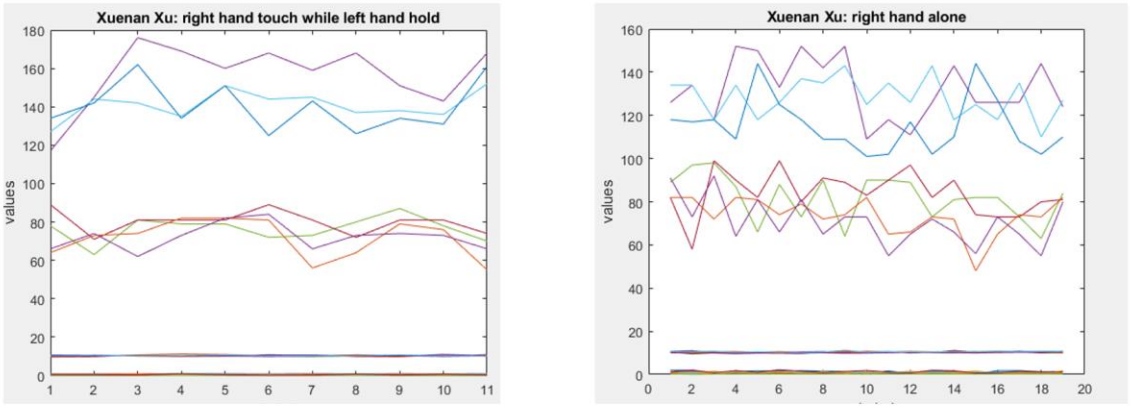

Fig. 2 The same user touch the screen by different position

These characteristics in different people showed a significant difference, as shown in Figure 3. Different people use the same posture, the value of the Y axis has a clear difference.

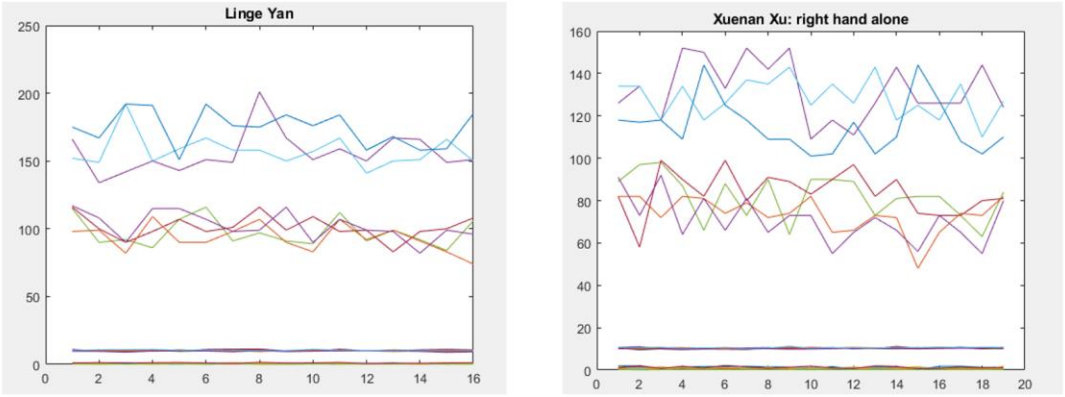

Fig. 3 The feature values of different person

In summary, this paper initially determine that these features meet the requirements of machine learning feature selection, suitable for training classifier. Whether the specific is really appropriate, but also through the results of the classifier and conclusions.

We collected a total of 4 different users of data. User 1 has 70 sets of data, user 2 has 47 sets of data, user 3 has 56 sets of data, and user 4 has 63 sets of data. A total of 236 sets of data, this

article will be the user as a mobile phone holder 1, the classification is defined as correct, others are regarded as wrong users.

3. Classification

1) One-class learning

One Class Learning means that your training data has only one type of positive (or negative) data, but not another type. At this moment, you need to learn the boundary of actually training data. And at this time cannot use the maximum margin, because you do not have two types of data. One-Class Learning is more suitable when only positive samples cannot define negative samples. Single-class algorithms only focus on the similarity or matching with samples, and do not make any conclusions about the unknown part. Many people say, "I do not know what love is, but I know what it is not love." When they blind date, they immediately check out their loved ones. They are a clever One-Class learner.

In fact, most of the security identification technology should be classified as One class learning. Because the initial state of the classifier contains only the user's own data. So in order to simulate the actual use, we conducted one class SVM modeling on the 70 correct data and carried out corresponding tests. The next 160 error data all identify correctly, there is no error.

2) Binary Classification

In order to better test whether the selected features are really appropriate, we still need to classify the data and test its accuracy. We used a variety of classifiers to train the data and test it, including: Naive Bayes, KNN, Logistic regression, Random forest, Decision tree, SVM. Accuracy using F1 for comparison, the formula is:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

And the result is as follows:

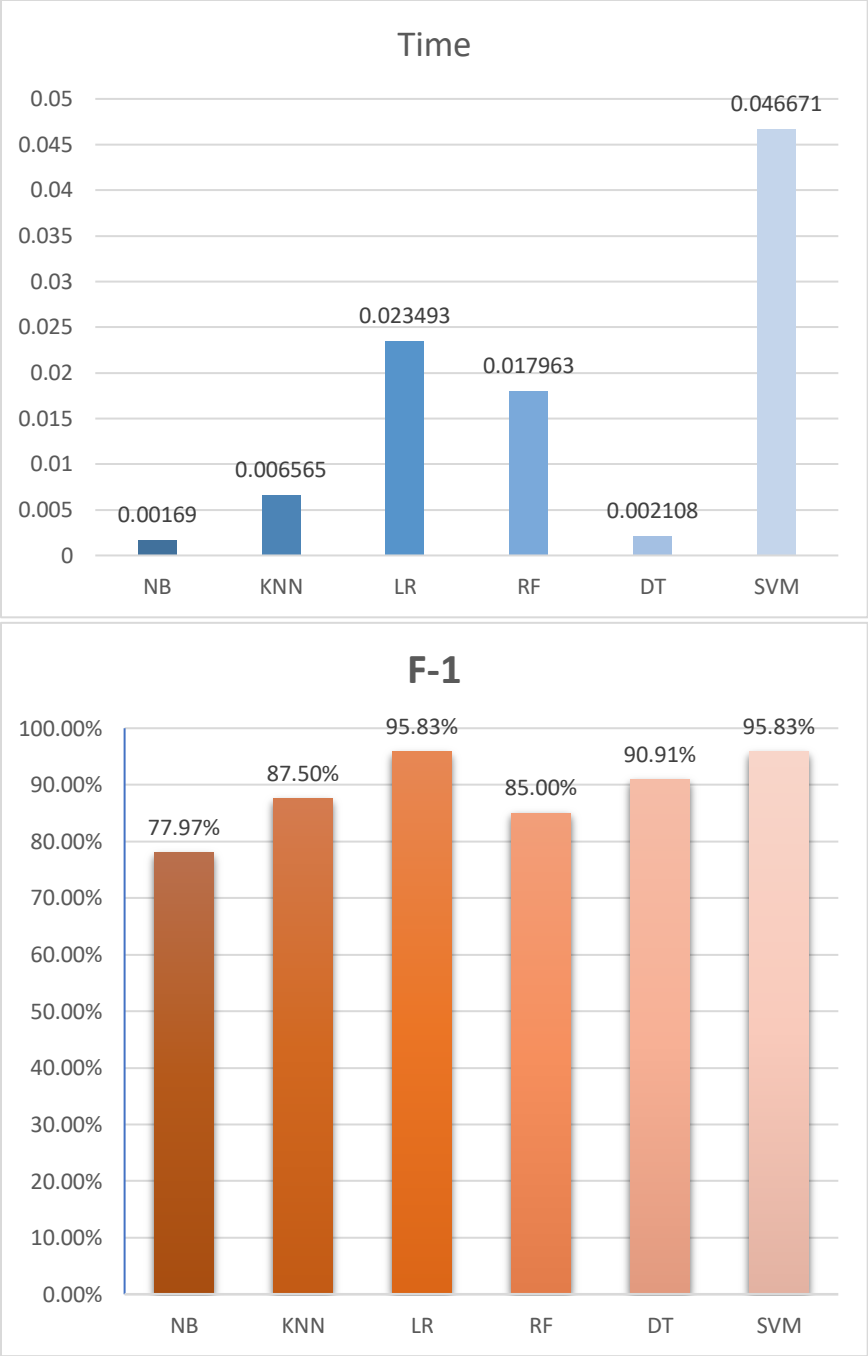

Fig. 4 Each classifier consumes time and accuracy

Through the chart we can clearly see that in all classifiers, the accuracy has remained at a high level. The SVM and logistic regression F1 value is reached more than 95%. All this shows that the features we select have very good differences and can be used as biometrics for security identification.

We then used grid search to optimize SVM and logistic regression training parameters. At the same time use cross-validation method to prevent over-fitting due to too little data. The result is still maintained at 90%.

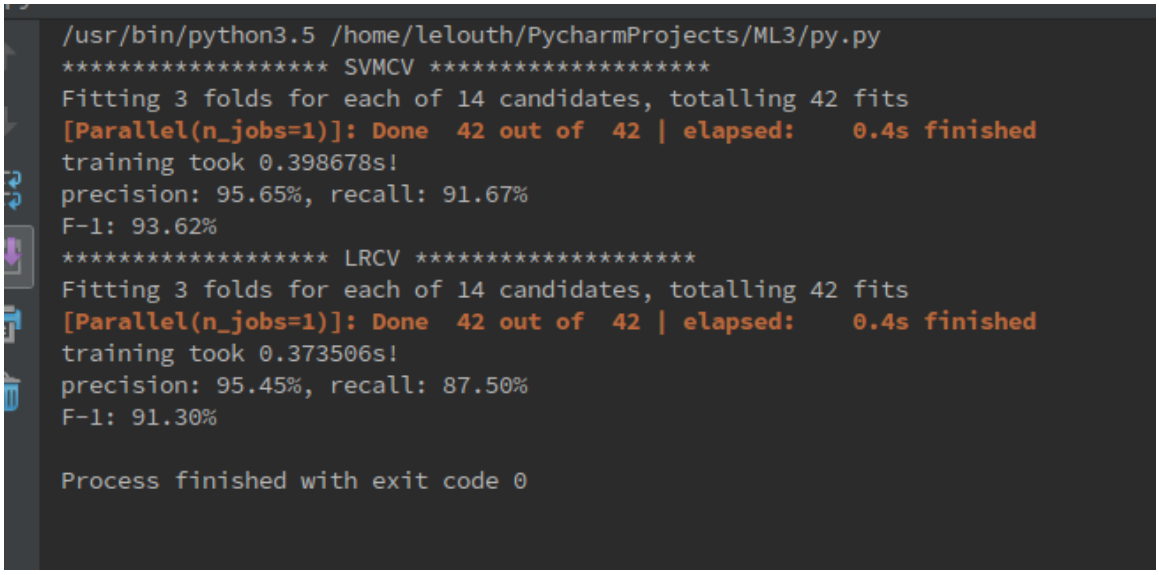

Fig.5 After optimizing the parameters, the results of logistic regression and SVM cross-validation

4. Conclusion

I. By introducing more features into the pre-authentication process, we could classify whether the “toucher” is a trusted user or an imposter who know the password.

II. The 4 features: acceleration, pressure, size and time interval, is appropriate and sufficient. We can build excellent classifier with these features.

III. Although the SVM and logistic regression cost more time than other classifier, these two way are better choice in this.

References

[1] Sharma, Vaibhav, and Richard Enbody. "User authentication and identification from user interface interactions on touch-enabled devices." *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 2017.

[2]Alzubaidi, Abdulaziz, and Jugal Kalita. "Authentication of smartphone users using behavioral biometrics." *IEEE Communications Surveys & Tutorials* 18.3 (2016): 1998-2026.

[3] Feng, Tao, et al. "Continuous mobile authentication using virtual key typing biometrics." *Trust, security and privacy in computing and communications (TrustCom), 2013 12th IEEE international conference on*. IEEE, 2013.

[4] Meng, Weizhi, et al. "Surveying the development of biometric user authentication on mobile phones." *IEEE Communications Surveys & Tutorials* 17.3 (2015): 1268-1293.

[5] Van Nguyen, Toan, Napa Sae-Bae, and Nasir Memon. "DRAW-A-PIN: Authentication using finger-drawn PIN on touch devices." *Computers & Security* 66 (2017): 115-128.