Zhen-yi XU, Yu KANG, Yang CAO, Yu-xiao YANG, 2019. Man-machine verification of mouse trajectory based on the random forest model. *Frontiers of Information Technology & Electronic Engineering*, 20(7):925-929.
https://doi.org/10.1631/FITEE.1700442

Front Inform Technol Electron Eng

Man-machine verification of mouse trajectory based on the random forest model

Key words: Man-machine verification; Random forest; Support vector machine; Logistic regression; Performance metrics

Corresponding author: Yu KANG
E-mail: kangduyu@ustc.edu.cn

Motivation

1. The identifying code is a promising technique for authorized access, and has been widely used in man-machine verification to prevent unauthorized access, theft of proprietary information, and insider net abuse.

2. In the behavioral verification application, the validation area is fixed, where some malicious applications can easily simulate the mouse drag action and complete the verification.

3. The challenge in engaging man-machine verification involves the correct classification of man and machine tracks.

4. A comprehensive movement analysis is needed to construct the trajectory features, which can help construct the classification model.

Front Inform Technol Electron Eng

Main idea

1. The comprehensive trajectory features are defined to create movement feature vectors, which represent the characteristics of the movement tracks.

2. The random forest (RF) classification model is adapted to handle the task of man-machine verification based on mouse trajectory data.

3. We compare the prediction performance of the proposed RF model with those of logistic regression (LR) and support vector machine (SVM) on both the training and testing data sets.

Front Inform Technol Electron Eng

Major results

1. The movement offset on the *y* axis can be chosen as the main features for trajectory classification.

Front Inform Technol Electron Eng

Table 3 Feature importance ranking

Index	Feature	Importance
0	t_{mean}	0.094 655
1	t_{var}	0.054 040
2	$v_{x\mathrm{mean}}$	0.048 928
3	$v_{x\mathrm{var}}$	0.083 988
4	x_{mean}	0.210 774
5	x_{var}	0.143 252
6	$v_{y\mathrm{mean}}$	0.041 525
7	$v_{y\mathrm{var}}$	0.078 242
8	y_{mean}	0.231 997
9	y_{var}	0.012 600

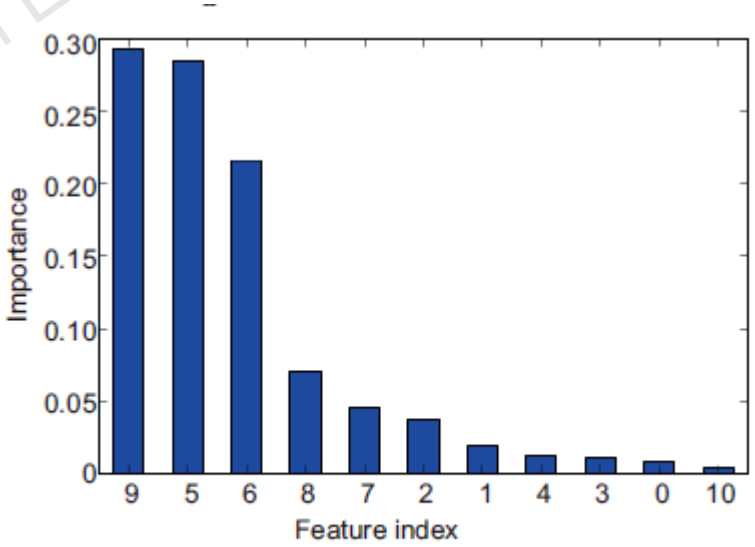

Fig. 3 Variable importance

Major results (Cont'd)

2. The RF model is superior to LR and SVM in terms of TNR, Precision, *F*-measure, *G*-mean, and w-Acc, but is worse than SVM in terms of TPR.

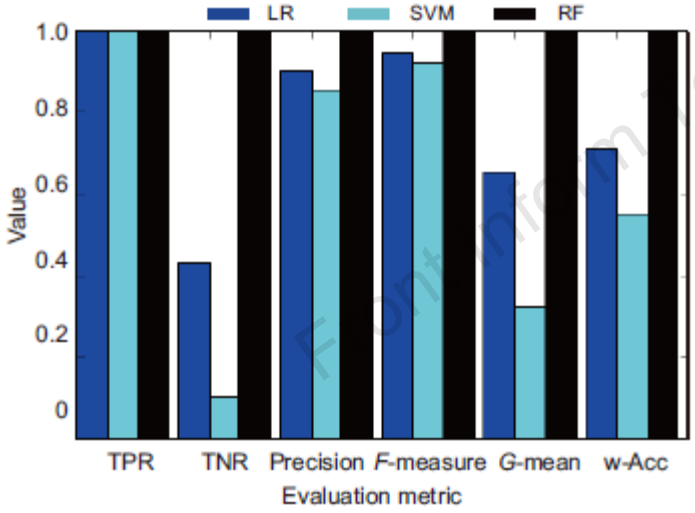

Fig. 4 Performance comparison
LR: logistic regression; SVM: support vector machine; RF: random forest; TPR: true positive rate; TNR: true negative rate; w-Acc: weighted accuracy

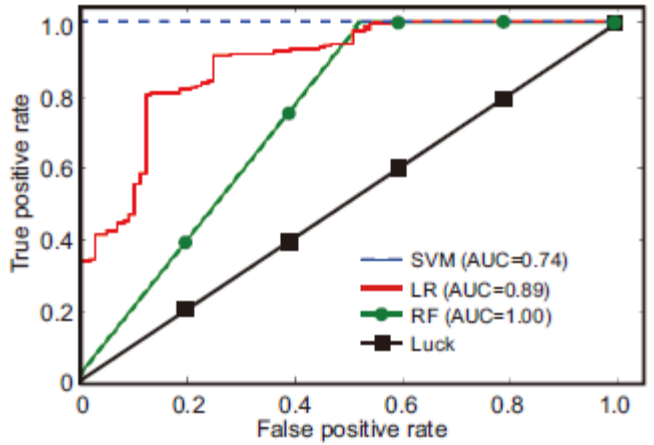

Fig. 5 ROC curve
LR: logistic regression; SVM: support vector machine; RF: random forest; ROC: receiver operating characteristic; AUC: area under curve

Conclusions

1. We constructed an RF model for man-machine verification based on the movement speed feature f_{speed}, movement offset feature f_{offset}, and movement interval feature f_{time} from the mouse movement trajectory dataset.

2. The comprehensive performance of the RF model is better than those of the SVM and LR models on the real-world mouse trajectory dataset, demonstrating the efficiency of our proposed method.

Front Inform Technol Electron Eng