Adaptive algorithms in accelerometer biometrics

Brazilian Conference on Intelligent Systems, 3th, 2014, São Carlos.
http://www.producao.usp.br/handle/BDPI/48587

Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo
Adaptive Algorithms in Accelerometer Biometrics

Paulo Henrique Pisani
Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Email: phpisani@icmc.usp.br

Ana Carolina Lorena
Universidade Federal de São Paulo
Departamento de Ciência e Tecnologia
Email: alore@unifesp.br

André C. P. L. F. de Carvalho
Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Email: andre@icmc.usp.br

Abstract—Nowadays, many services are available from mobile devices, like smartphones. A growing number of people are using these devices to access bank accounts, social networks and to store personal information. However, common authentication mechanisms already present in these devices may not provide enough security. Recently, a new authentication method, named accelerometer biometrics, has been proposed. This method allows identification of users using accelerometer data. Accelerometers, usually present in modern smartphones, are devices that measure acceleration forces. In accelerometer biometrics, a model is induced for the user of the smartphone. However, as a behavioral biometric technology, user models may become outdated over time. This paper investigates the use of adaptation mechanisms to update biometric user models induced by accelerometer data along the time. The paper also proposes and evaluates a new adaptation mechanism with promising experimental results.

I. INTRODUCTION

Several services, both personal and corporate, are currently available in the Internet. Many of these services can be accessed from mobile devices, such as tablets and smartphones. A recent study [1] showed that worldwide smartphone sales reached the sum of 225 million units just in the second quarter of 2013. In fact, an ever-increasing number of people is now using these devices to manage bank accounts, access corporate systems, use social networks and store personal information. All of these tasks involve dealing with sensitive information. However, does commonly used authentication mechanisms provide enough security to them?

A study on security of mobile devices [2] showed a worrying number: only 13% of the participants used PIN (Personal Identification Number) or visual code (an authentication method present in Android devices). The main reason given was that, without authentication, it is faster to use the device.

Recently, a new authentication mechanism based on accelerometer data was proposed [3]. This technology is called accelerometer biometrics [4] and can use data already present in the smartphone accelerometer. Accelerometer biometrics can be used to authenticate users without the need of interrupting their activities to enter a password, for instance. As a result, more users can feel comfortable to employ authentication mechanisms in their mobile devices.

There are not many studies on accelerometer biometrics using smartphone data. As a consequence, there are some open questions, like: does user behaviour changes over time (concept drift) on accelerometer biometrics using smartphone data? If so, how does it affect user recognition performance? In data stream mining, concept drift refers to the fact that the distribution which generates data is non-stationary [5]. In a previous work [6], we studied another biometric technology and concluded that the adaptation to concept drift has a key impact on user recognition performance. However, it is unclear how it could affect accelerometer biometrics in a data stream context. Several studies found in the literature analysed this technology inducing a static user model only. This paper investigates the user recognition performance over time using accelerometer data, considering a data stream context. Moreover, modifications made to one class SVM algorithm and a previous adaptive algorithm proposed by the authors, which can deal with concept drift, are also presented and investigated.

The next sections of this paper are organized as follows: Section II introduces the basic concepts of accelerometer biometrics; Section III presents the Self-Detector algorithm, some current adaptive versions and the new adaptive proposal; Section IV details the experimental setup, such as datasets, extraction features, evaluation methodology and parameter values; Section V shows the obtained results; and Section VI presents the main conclusions from this study.

II. ACCELEROMETER BIOMETRICS

Accelerometer biometrics has the goal of recognizing users by accelerometer data. The term accelerometer biometrics was used in a recent competition by Kaggle [4]. This work adopts the same term to define this technology. In the literature, accelerometer biometrics may also be referred to as cell phone-based biometrics [7].

This work focuses on accelerometer biometrics using data from mobile devices, like smartphones. One of the first studies to investigate the use of smartphone accelerometer data was [3], which considered users walking at three different speeds. Afterwards, [7] showed that other activities, like walking, jogging, ascending and descending stairs, can be used to recognize users by their smartphone accelerometer data. In [8], the authors evaluated three classification algorithms to perform this task: Hidden Markov Model (HMM), Support Vector Machines (SVMs),
and k-Nearest Neighbour (kNN). According to their tests, kNN obtained the best predictive performance. A recent work [9] also proposed a new distance metric, called Cross Dynamic Time Warping Metric, which achieved the best performance in the experiments.

A technology related to accelerometer biometrics is gait biometrics, which usually involves recognizing users by their walking pattern [10]. Data for gait biometrics can be obtained from different sources: visual data or sensors (e.g. accelerometer). As gait biometrics is a behavioural technology, it is expected that it is subject to concept drift. The work of [11] claims to be the first in literature to study the issue of time in gait biometrics. They used visual data in their experiments. According to them, users can still be recognized over 9 months using a static model. Their work did not study, however, the use of adaptive classification algorithms.

This paper adopts a different approach by using accelerometer data captured from smartphones and studying the performance of adaptive classification algorithms over time.

III. SELF-DETECTOR

This section describes the Self-Detector algorithm, some current adaptive versions and a new adaptive proposal. This classification algorithm was chosen due to its good predictive performance on a previous work in biometrics [6]. Furthermore, since it is instance-based, its model adaptation is simpler to perform, as argued by [12]. Additionally, another instance-based algorithm, kNN, obtained the best predictive performance for accelerometer data in a previous study [8].

A. Standard Non-adaptive Version

Self-Detector is an instance-based algorithm part of the positive selection class from immune algorithms. The standard Self-Detector [13] stores training examples from a user as detectors and assign a constant radius to each of them. When an input example is presented to the algorithm, all detectors are compared to it. If any detector matches the example, it is classified as self (legitimate user) and, otherwise, as non-self (intruder). In this paper, a detector matches a example if the distance between its center and the example is smaller than its radius. The original version of this algorithm uses a ROC analysis to define the self radius. A different approach is used here, as defined in Section IV.

B. Current Adaptive Versions

Current adaptive Self-Detector versions work by applying an updating procedure when a new example is classified as positive by the algorithm. Here, we discuss three adaptation versions for Self-Detector: Growing, Sliding and Usage Control. Growing and Sliding are based on [14] and [15], which used similar ideas. These approaches were previously used for Self-Detector in [6]. In the Growing method, each example classified as positive (from the legitimate user) is included as a new detector in the detector set. The Sliding works similarly, but it also removes the oldest detector from the detector set when a new detector is added. This makes the amount of detectors constant in Sliding and, therefore, more memory efficient than Growing, which only grows the detector set. It is important to note that an example misclassified as positive will never be removed in Growing.

The third version, Usage Control, introduced by our previous work in [6], assesses which detectors are more used in order to preserve them. It is based on the idea that most used detectors better represent the current user behaviour. The storage of examples in memory and their replacement according to their usage was also discussed in the context of biometrics in a technical report [16], although their approach is different from Usage Control. Two additional attributes are assigned to each detector in Usage Control:

- **Usage count**: increases every time the detector matches an example.
- **Recent usage**: decreases every time another detector matches an example. If a detector matches an example, it returns to a maximum value (here we empirically adopted 10, same value of [6]). The detector also assumes the maximum value when it is first generated.

In Usage Control, when a new example is presented, detectors are checked from the oldest to the newest. If a detector matches the example, the two additional attributes are updated. All detectors with Recent usage equals to zero are ordered by Usage count. The detector with the lowest Usage count value is removed and a new detector is added to the detector set using the matched example. These two additional attributes enables the removal of detectors with low usage without removing new detectors instantly (as their Usage count is zero when they are created). If there is no detector with Recent usage equals to zero, no adaptation occurs and the recognized example is discarded.

C. New Adaptive Version: Usage Control S

In the previous version of Usage Control (as well as other adaptive Self-Detectors), an example being recognized as positive is usually enough to be included as a new detector (as described in the last section). Consequently, a misclassified example can be wrongly included in the detector set, increasing false acceptance rates. In view of this problem, we propose to use an example as a new detector only if more than one detector matched it. This new rule assumes that an example matched by two or more detectors has a higher level of confidence that it is a true positive. Conversely, examples matched by only one example have low level of confidence and, therefore, are not used as a new detector.

Another modification is the increase of the value of the attribute Usage count for all detectors that match the input example (Recent usage decreases for all other detectors, which do not match the input example). The first version of Usage Control only increased the Usage count of the first detector to match the input example.
Table I: Results in Datasets A and B. Values between parenthesis are the standard deviation for each measure.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Dataset A</th>
<th>Dataset B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FAR (0.004)</td>
<td>FRR (0.000)</td>
</tr>
<tr>
<td>One-class SVM (No adaptation)</td>
<td>0.141</td>
<td>0.518</td>
</tr>
<tr>
<td>One-class SVM (Growing)</td>
<td>0.140</td>
<td>0.518</td>
</tr>
<tr>
<td>One-class SVM (Sliding)</td>
<td>0.203</td>
<td>0.192</td>
</tr>
<tr>
<td>Self-Detector (No adaptation)</td>
<td>0.177</td>
<td>0.266</td>
</tr>
<tr>
<td>Self-Detector (Growing)</td>
<td>0.387</td>
<td>0.113</td>
</tr>
<tr>
<td>Self-Detector (Sliding)</td>
<td>0.203</td>
<td>0.192</td>
</tr>
<tr>
<td>Self-Detector (Usage Control)</td>
<td>0.267</td>
<td>0.145</td>
</tr>
<tr>
<td>Self-Detector (Usage Control S)</td>
<td>0.148</td>
<td>0.219</td>
</tr>
</tbody>
</table>

As a consequence, other detectors that could also match the example are considered as not used. Thereby, these detectors could be removed even though they can represent well the current user behaviour. The new version of Usage Control, which contains the modifications described in this section, is named Usage Control S.

IV. EXPERIMENTAL SETUP

A. Datasets and extracted features

Performance was assessed using two datasets from WISDM, which has data captured from smartphone accelerometer. These datasets are Activity Prediction [17] and Actitracker [18]. The newest versions of these datasets were used, as available in http://www.cis.fordham.edu/wisdm/dataset.php.

As the name suggests, accelerometer measures acceleration forces, usually at a given frequency. In these datasets, accelerometers provided accelerations for the x, y and z axis. Before the experiments, these series of data had to be preprocessed. A common strategy is to divide the series into windows of a predefined length. Afterwards, features are extracted from these windows. This paper considers a window equivalent to 2s of data, as in [19]. Since in the datasets used, data was sampled at 20Hz, every window corresponds to 40 measures for each of the three axes. To divide the 40-example windows, this work used only the order of the examples per user, ignoring timestamps.

According to [19], which compared several different feature vectors from accelerometer data, the best results were obtained by using features in the frequency domain, such as by using magnitudes resulted from a Fast Fourier Transform (FFT). Hence, this work applied FFT over each of the 40-example window and used the magnitude of the first five components of each axis to generate the feature vector, as in [19]. Hence, the feature vector is composed by 15 features (5 for each axis). The original datasets have data for six activities, but this study considered only the “walking” activity. In both datasets, the majority of examples (40%) corresponds to this activity. A summary of the processed datasets is shown next (users with less than 100 examples were discarded):

- Activity Prediction (Dataset A): 36 users and a total of 10,591 examples;
- Actitracker (Dataset B): 131 users and a total of 29,190 examples.

Based on the preprocessed data, a data stream is generated for each user. An approach similar to [6] is used, in which streams are formed by all examples from the legitimate user interleaved with examples from other users randomly chosen (intruders). The data stream has a fixed rate of 30% of negative examples, as in [6]. The order of the examples in the dataset is preserved, in order to allow the identification of concept drift.

B. Evaluation methodology

This study considers the user recognition task as a one-class classification problem. Hence, only examples from the legitimate users (one class) are used to generate the user model. Classification models are induced by the learning algorithm using the first \( N \) positive examples from the user (\( N \) assumed the value of 40, as in [6]). Afterwards, in the test phase, each example in the data stream is presented to the algorithm, for classification and model adaptation. No class label is provided to the algorithm, thus, the adaptation does not use this information. When generating a data stream for a given positive user, examples already used in the training phase are not present in the user data stream for the test phase.

Data streams are generated per user and, therefore, the results are obtained per user. The values reported here are the average considering all users. Moreover, due to the stochastic nature of the data stream generation (negative examples are interleaved randomly), all experiments are repeated 30 times and their average values are reported. Since the data stream is imbalanced, balanced accuracy was used to report accuracy.

C. Classification algorithms and parameters

All Self-Detector variations described earlier are used in the experiments. Parameter optimization was performed by the same method used in [20]. In the case of Self-Detector, the optimization was performed in the non-adaptive version and the same parameter values were used for the adaptive counterparts. Self-radius assumed the value of 0.01 in both datasets (using cosine distance).

Besides Self-Detectors, the One-class Support Vector Machine (OCSVM) [21] algorithm is used in the experiments (implementation of LIBSVM [22]). The same parameter optimization used for Self-Detectors was applied for OCSVM, with \( \epsilon = 10^{-6} \) and RBF kernel. \( \nu \) assumed 10\(^{-4} \) in both datasets, \( \gamma \) assumed 10\(^{-5} \) in Activity Prediction and 10\(^{-4} \) in Actitracker (tested values were \( \nu = [10^{-1}; 10^{-2}; 10^{-3}] \) and \( \gamma = [10^{0}; 10^{-1}; 10^{-2}; 10^{-3}; 10^{-4}; 10^{-5}; 10^{-6}] \)).

All algorithms used the raw dataset after the application of FFT. LIBSVM recommends that feature vectors are
rescaled to [0;1] or [-1;1] [23]. However, results of SVM using min-max recale [24] were on average worse for accelerometer data (rescale factors were obtained using the training set per user). A possible explanation for this effect is that all extracted features are FFT magnitudes and min-max may have changed relations between features.

In addition to the static OCSVM, two adaptive approaches using ideas from [15] and [14] were included in this study: OCSVM (Growing) and OCSVM (Sliding). They work similarly to the Self-Detector Growing and Sliding versions. In Growing, the initial training set is stored. Whenever an example is recognized as positive by OCSVM, it is added to the stored training set and OCSVM is retrained to induce a new model. In Sliding, the idea is the same, but the oldest example from the stored set is removed when a new example is added.

V. EXPERIMENTAL RESULTS

A. Global results

Table I shows the overall performance values for both datasets regarding false acceptance rate (FAR), false rejection rate (FRR) and accuracy rate (balanced). According to the accuracy results, Self-Detector adaptation methods obtained higher performance than non-adaptive algorithms in most of the cases. Usage Control S, proposed here, consistently reached the best accuracy for both datasets. This suggests that including only examples recognized by more than one detector (with higher level of classification confidence) improves the performance obtained by the previous version of Usage Control.

All adaptive methods decreased their FRR when compared to the standard versions, with the exception of OCSVM. This can indicate the occurrence of concept drift. Thus, the adaptation of the user model can improve the predictive performance. The high FRRs by OCSVM may have negatively impacted its adaptive versions, what can result in less examples for model adaptation. OCSVM Sliding, for instance, presented accuracy worse than the non-adaptive case. The lowest FAR for OCSVM may have contributed to further reduce this rate in the adaptive versions.

Self-Detector Growing also obtained accuracy lower than the non-adaptive Self-Detector. As Growing only increases its detector set, the range of acceptable patterns also rise, explaining the highest FAR and lowest FRR observed in the results. However, this resulted in lower accuracy performance.

A Friedman statistical test [25] showed that there are significant differences among the algorithms for p < 0.10. Among the Self-Detector versions, Usage Control S obtained the best FAR values. This was expected, since more reliable data are used in the adaptation of this model. Furthermore, this algorithm improved all rates over the non-adaptive Self-Detector. Regarding FAR, OCSVM obtained best values, but at the cost of highest FRR.

B. Performance over time

The FAR and the FRR were also measured over time. For such, the data stream of each user was divided into blocks and the average performance was measured. As data streams for the Dataset A are on average longer, the streams were divided into three blocks for the Dataset A and in two for the Dataset B. Figures 1 and 2 show FRR and FAR measured through the stream. Since they presented accuracy lower than their non-adaptive counterparts, Self-Detector (Growing) and OCSVM (Sliding) were not considered in this section.

These graphs show that the FRR increased over time for all algorithms, including the adaptive versions. Thus, the user model adaptation to the current user behaviour needs to be improved. Regarding the FAR, Usage Control (first version) and Sliding obtained the worst performance. Compared to the other algorithms, they had a higher tendency to increase this rate over time. The more rigorous adaptation mechanism of Usage Control S improved the FAR. This indicates that misclassified examples could be degrading the classification performance of the first version of Usage Control.

C. Analysis of adaptation

The maximum correlation between detectors and the positive examples through the data stream can indicate if the model was successfully adapted over time. Figure 3 shows, for four algorithms, a plot for the maximum correlation among all detectors and positive examples through the data stream. This section only analyses Self-Detector, which is based on the concept of detector set
(Growing was not considered in this section due to its lower global accuracy). Each graph corresponds to the data stream from one user in order to highlight the effect of model adaptation. Three users with different behaviours were selected from the datasets to show how the adaptive algorithms perform on diverse cases. Users 1 and 2 are from Dataset A and User 3 from Dataset B.

In these graphs, higher values are better, since they represent higher correlations between the detector set and legitimate user examples. According to our analysis, concept drift does not strongly affect all users. User 1 is one example. However, the adaptive approaches have not harmed the user model over time. In the no adaptation scenario, the maximum correlation steadily decreases for User 2 and 3, showing that the user model may have become outdated and, therefore, does not represent the current user behaviour. This suggests that concept drift occurs for these two users.

All adaptation methods kept the correlation higher by automatically updating their set of detectors. For User 3, however, Usage Control S kept correlation values lower than the other adaptive methods. As the no adaptation graphs shows, this user had a strong behaviour change, making the model adaptation more difficult. Even though, Usage Control S obtained higher values compared to the non-adaptive algorithm.

This relates to a possible reason for the FRR increase over time, even for the adaptive approaches. Adaptive approaches used here can only adapt to incremental drifts, as it is very difficult, without knowing the true labels, to relate abrupt changes as being a behaviour change and not examples from intruders. As a result, models for users with such a behaviour were not properly updated to these changes, contributing to an increased FRR through the stream.

VI. CONCLUSION

This work investigated accelerometer biometrics in a data stream context, considering that concept drift can occur. A modification over a previous adaptive algorithm was also proposed and evaluated. According to experimental results, the proposal obtained better accuracy among all tested algorithms.

Overall, the predictive results are promising, as they correspond to a relative small window of data. A larger window may improve the predictive performance of all algorithms. This work recommends the use of accelerometer biometrics as an additional layer of security. A FAR of 15% to 20% may be good for many applications, but not for others, like banking applications. However, the combination of accelerometer biometrics with other authentication methods can enable the use of this technology in scenarios which demand higher levels of security.

The analysis conducted in this paper suggests that concept drift occurs in accelerometer biometrics data, but not for all users. Additionally, Self-Detector obtained better accuracy performance than OCSVM in the configuration adopted here. Usage Control S improved all rates over the non-adaptive Self-Detector, indicating that the more rigorous adaptive method is suitable for accelerometer biometrics using smartphone data.

In future work, other scenarios can be considered, such as different rates of positive/negative examples and different window sizes. Other combinations of extracted features can also be investigated, like combining time and frequency domain features.

ACKNOWLEDGMENT

The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for financial support.

REFERENCES

Figure 3: Effect of adaptation methods on the positive examples of different users.


