Biometrics in a data stream context

Paulo Henrique Pisani[1], **André C. P. L. F. de Carvalho**[1]

[1]Universidade de São Paulo (USP)
Instituto de Ciências Matemáticas e de Computação (ICMC)
São Carlos – SP – Brazil

paulo.pisani@ufabc.edu.br, andre@icmc.usp.br

Abstract. *Biometric systems can provide safer authentication. However, biometric features may change over time, impacting the recognition performance due to outdated biometric references. It raises the need to automatically adapt the references over time, by using adaptive biometric systems. This thesis studied several aspects of adaptive biometric systems in a data stream context. Based on this investigation, it was observed that the best choice for each aspect can be user dependent. This motivated the proposal of a modular adaptive biometric system, which can select a different configuration for each user. It also generalizes several baselines and proposals into a single modular framework, while opening numerous opportunities for future work.*

1. Introduction

In view of the increasing use of digital identities, concerns regarding data exposure due to identity theft have increased. These concerns emphasize the need of enhanced authentication mechanisms. The commonly used login and password combination may not provide enough security for some applications, since a password can be easily copied and even guessed when weak passwords are used. Using a card or token can improve the security of a system, although the possession of the device may be compromised.

Biometrics is an alternative to the use of passwords to perform authentication. *Biometrics* is defined as the automatic recognition of individuals through their biological or behavioural traits [Jain et al. 2016]. A number of attributes have been used in the literature for this purpose, such as fingerprint, iris, keystroke dynamics, gait, etc. Several requirements must be met by biometric features to be used in a biometric system, such as universality, distinctiveness, collectability and permanence [Jain et al. 2004]. Permanence refers to the fact that the biometric features will remain stable (not change) over time. Nonetheless, recent studies have shown that biometric features extracted from several traits may change over time [Roli et al. 2008]. Consequently, the biometric reference may not reflect the current user data anymore, an issue known as *template ageing* [Jain et al. 2016]. As a result, the recognition performance of biometric systems may be affected.

The thesis [Pisani 2017] dealt with *biometrics in a data stream context*. In Data Stream Mining, a data stream is defined as a potentially unlimited sequence of examples that are accessed sequentially [Aggarwal et al. 2003]. In the thesis, a data stream is a sequence of biometric samples received by the biometric system. Data distribution from the genuine user is subject to changes over time, also referred to as *concept drift* [Žliobaitė 2010]. In this context, the goal is to keep the biometric reference updated to

changes observed in the data stream over time. Performing periodical enrollment sessions for all users in the system is a straightforward method to update the biometric reference. However, this procedure can imply in high operational costs and user annoyance. An alternative is the adoption of an adaptive biometric system, which is able to automatically adapt the biometric reference over time [Rattani 2015]. Adaptive biometric systems are the focus of the thesis.

Several aspects of adaptive biometric systems were discussed throughout the thesis. In the end, a modular adaptive biometric system was proposed, able to choose different aspects of adaptive biometric systems per user. Next sections are organized as follows: Section 2 states the problem, discusses the studied hypotheses and presents the main objective of the thesis; Section 3 briefly describes the main contributions of the thesis and researchers which worked in collaboration during the PhD; and, Section 4 enumerates the publications obtained from the research of the thesis and related work performed later.

2. Problem, hypotheses and objective

The field of adaptive biometric systems is relatively new and there are still several open questions. This fact, combined to recent work reporting improvements of adaptive biometrics over non-adaptive approaches [Giot et al. 2012a, Rattani 2015], highlights the importance of new studies on adaptive biometric systems. Additionally, behavioural modalities are subject to a larger number of changes in short periods than physical modalities [Giot et al. 2012b]. Nevertheless, most work has been concerned with physical modalities, like fingerprint and face. The thesis aimed to reduce this gap by focusing on *behavioural modalities*.

Problem: the problem investigated in this thesis is the ***design of adaptive biometric systems for behavioural modalities in a data stream context***. A biometric system in such a context needs to consider that it will receive biometric samples sequentially and that the biometric features may change over time. Furthermore, these changes can occur in a short time span for the behavioural modalities.

Among the behavioural modalities, the thesis focused on keystroke dynamics [Pisani and Lorena 2013] and accelerometer-based gait biometrics / accelerometer biometrics [Kaggle 2013]. They were chosen due to the availability of suitable datasets. To study adaptation, a dataset needs to contain several samples for each user and ideally be acquired at different sessions. Moreover, most of the work in the field have dealt with physical modalities, indicating an opportunity for new research on behavioural modalities. These modalities usually imply in lower discriminative power than physical modalities (e.g. face, fingerprint, iris), introducing additional challenges to the biometric system. Furthermore, they tend to be subject to faster changes over time than physical ones, like face, for example [Giot et al. 2012b], emphasizing the importance of additional studies for behavioural biometric modalities in a data stream context.

There are many aspects concerning the design of adaptive biometric systems, such as adaptation strategy, combination of genuine and impostor models, score normalization and choice of adaptation rules per user via a modular framework. They were all evaluated in the thesis and briefly described next, along with the investigated hypotheses.

The immune-based *Self-Detector* algorithm was successfully applied to keystroke dynamics, but in a scenario that did not consider biometric reference adaptation

[Pisani and Lorena 2015]. Even though, that work suggested that the typing rhythm changes over time. However, is it possible to turn *Self-Detector* into an adaptive algorithm? *Self-Detector* is part of the positive selection of immune algorithms [Stibor and Timmis 2005]. Adaptation could be performed by changing the detector set to control its usage for matching, in order to keep only the most used detectors, leading to the first hypothesis. The research presented in the thesis proposed some alternatives in this line, with some versions of *Usage Control*.

- **H1**: Positive selection can become an adaptive class of immune algorithms by controlling the usage of the detectors, enabling its use for biometrics in a data stream context.

To the best of the author's knowledge, current adaptive biometric systems only adapt a model for the genuine user in the biometric reference. As a result, samples classified as impostor are discarded and they could contain important information. This brings the following question: would the use of an additional impostor model generated from samples classified as impostor be able to enhance the recognition performance? The second hypothesis comes from this reasoning.

- **H2**: The use of a negative/impostor model, in addition to the positive/genuine model, can improve the recognition performance of adaptive biometric systems.

Furthermore, studies in biometrics have applied score normalization [Poh et al. 2009] to refine the final classification decision. Score normalization procedures often increase the class separability, which results in improved performance of biometric systems. However, to the best of the author's knowledge, score normalization has not been used for biometrics in a data stream context. Hence, could score normalization be used to increase adaptive biometric systems performance by a better refinement of the final decision? It means that by using score normalization, a more effective threshold could also be chosen, which can in turn increase the amount of correctly accepted genuine samples for adaptation, while reducing the amount of impostor samples wrongly accepted for adaptation. It results in the third hypothesis.

- **H3**: Score normalization can improve the recognition performance of adaptive biometric systems.

During the evaluations performed for the previous hypotheses, it was observed that the best performing adaptation strategy for a dataset is not always the best strategy for another dataset [Pisani et al. 2015]. This may be caused by distinct *change patterns* among the datasets. A *change pattern* is defined here as the way that the biometric features change over time. Following this reasoning, the work from [Pisani et al. 2015] has reported that the typing rhythm change can differ depending on the user. Moreover, the work from [Poh et al. 2015] showed that the recognition performance changes in different ways for each user over time. Another study has also suggested that different groups of users may need modifications in the adaptation strategy and parameters [Rattani et al. 2009].

By analyzing these findings, the following question was raised: should a different adaptation strategy be chosen for each user in the same dataset to improve the recognition performance? A system able to select the most suitable adaptation strategy for each user would bring adaptive biometric systems to another level: the choice of adaptation rules per user. To the best of the author's knowledge, an adaptive biometric system with these capabilities has never been proposed. It motivates the next hypothesis.

- **H4**: An adaptive biometric system can be divided into modules, such that it can choose different module implementations for each user. Therefore, distinct adaptation rules can be assigned to each user.

The thesis presented and investigated the hypothesis that this new adaptive biometric system can be designed following a modular framework, in which each aspect of the system is divided into modules. As a result, by choosing the implementations of the modules, the adaptive behaviour of several current adaptive biometric systems could be reproduced. Within this framework, the selection of adaptation rules per user is equivalent to choosing the implementations of each module per user. Since the best adaptation strategy may differ among users, this modular system would also potentially be able to obtain the best recognition performance on all datasets.

Objective: the main objective of the thesis was to ***enhance the design of adaptive biometric systems for behavioural modalities in a data stream context to improve the recognition performance of these systems***.

This is accomplished by proposing new adaptation strategies, using an additional impostor model, applying score normalization and selecting adaptation rules per user. In the end, a framework for a *modular adaptive biometric system* was proposed, which can generalize several baseline systems and, in the future, include all aspects of an adaptive biometric system investigated in the thesis. This modular system is able to select adaptation rules per user, bringing adaptation to a new level.

3. Main contributions

The first contribution of the thesis was the evaluation methodology for biometrics in a data stream context. Different methodologies have been adopted in previous work, however, they are not entirely suitable to properly evaluate a biometric data stream. The experiments carried out in the thesis followed the new evaluation methodology.

Throughout the thesis, several aspects to enhance the design of adaptive biometric systems for behavioural modalities in a data stream context were discussed: *Usage Control* versions for the immune-based *Self-Detector*, combinations of genuine and impostor models in the *Enhanced Template Update* framework and application of score normalization for biometrics in a data stream context. Each of these aspects resulted in important contributions, which led to several journal publications, as shown in the next section.

Based on the investigation of the diverse aspects of adaptive biometric systems listed in the previous paragraph, it was noted that the best choice for each of them can be user dependent. Consequently, the modular adaptive biometric system, named *ModBioS*, was proposed. *ModBioS* can generalize several baselines and proposals into a single modular framework. Different from proposals in the literature, *ModBioS* can select an adaptation strategy per user, representing a key contribution from the thesis.

Moreover, during the development of the PhD research, several researchers worked in collaboration. It included collaborations from Brazil and abroad:

- Prof. Ana Carolina Lorena (UNIFESP/Brazil);
- Prof. Romain Giot (Université de Bordeaux/France);
- Prof. Norman Poh (University of Surrey/England).

The research from this thesis opened a number of opportunities for future work, as described in Chapter 8 of the thesis. Currently, the author is an Assistant Professor (Professor Adjunto A) at Universidade Federal do ABC, where he plan to further extend the current research on biometrics and the collaborations established during the PhD.

4. Publications

Scientific papers have been published during the development of the research for this thesis. The papers that were published and are related to the thesis are presented next:

• [$Journal - Qualis$ **A2**] Pisani, P. H.; Poh, N.; Carvalho, A. C. P. L. F.; Lorena, A. C. Score Normalization applied to Adaptive Biometric Systems. Computers & Security, Elsevier, v. 70, p. 565-580, 2017. DOI: 10.1016/j.cose.2017.07.014

• [$Journal - Qualis$ **B1**] Pisani, P. H.; Lorena, A. C.; Carvalho, A. C. P. L. F. Adaptive algorithms applied to accelerometer biometrics in a data stream context. Intelligent Data Analysis, IOS Press, v. 21, n. 2, p. 353-370, 2017. DOI: 10.3233/IDA-150403

• [$Journal - Qualis$ **A2**] Pisani, P. H.; Giot, R.; Carvalho, A. C. P. L. F.; Lorena, A. C. Enhanced template update: Application to keystroke dynamics. Computers & Security, Elsevier, v. 60, p. 134-153, 2016. DOI: 10.1016/j.cose.2016.04.004

• [$Conference - Qualis$ **B2**] Pisani, P. H.; Lorena, A. C.; Carvalho, A. C. P. L. F. Ensemble of adaptive algorithms for keystroke dynamics. In: 2015 Brazilian Conference on Intelligent Systems (BRACIS). p. 310-315, IEEE, 2015. DOI: 10.1109/BRACIS.2015.29

• [$Conference - Qualis$ **A1**] Pisani, P. H.; Lorena, A. C.; Carvalho, A. C. P. L. F. Adaptive approaches for keystroke dynamics. In: 2015 International Joint Conference on Neural Networks (IJCNN). p. 1-8, IEEE, 2015. DOI: 10.1109/IJCNN.2015.7280467

• [$Journal - Qualis$ **B1**] Pisani, P. H.; Lorena, A. C.; Carvalho, A. C. P. L. F. Adaptive positive selection for keystroke dynamics. Journal of Intelligent & Robotic Systems, Springer, v. 80, n. 1, p. 277-293, 2015. DOI: 10.1007/s10846-014-0148-0

• [$Conference - Qualis$ **B2**] Pisani, P. H.; Lorena, A. C.; Carvalho, A. C. P. L. F. Adaptive algorithms in accelerometer biometrics. In: 2014 Brazilian Conference on Intelligent Systems (BRACIS). p. 336-341, IEEE, 2014. DOI: 10.1109/BRACIS.2014.67

• [$Journal - Qualis$ **B1**] Pisani, P. H.; Lorena, A. C. A systematic review on keystroke dynamics. Journal of the Brazilian Computer Society, Springer, v. 19, n. 4, p. 573-587, 2013. DOI: 10.1007/s13173-013-0117-7

• [$Conference - Qualis$ **B4**] Pisani, P. H.; Lorena, A. C.; Carvalho, A. C. P. L. F. Algoritmos imunológicos adaptativos em dinâmica da digitação: um contexto de fluxo de dados. In: Anais do X Encontro Nacional de Inteligência Artificial e Computacional - ENIAC. p. 1-12, 2013.

The research from the paper "Ensemble of adaptive algorithms for keystroke dynamics", published during the PhD, was extended during a Post-Doctorate at ICMC-USP, where the following publication was obtained:

• [$Journal - Qualis$ **A1** $- Accepted$] Pisani, P. H.; Lorena, A. C.; Carvalho, A. C. P. L. F. Adaptive Biometric Systems using Ensembles. IEEE Intelligent Systems, 2018. DOI: 10.1109/MIS.2018.111144956

Acknowledgments

The authors would like to thank Brazilian agencies CAPES and FAPESP 2012/25032-0 for their financial support.

References

Aggarwal, C. C., Han, J., Wang, J., and Yu, P. S. (2003). A framework for clustering evolving data streams. In *Proceedings of the 29th International Conference on Very Large Data Bases - Volume 29*, VLDB '03, pages 81–92. VLDB Endowment.

Giot, R., Rosenberger, C., and Dorizzi, B. (2012a). Hybrid template update system for unimodal biometric systems. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE.

Giot, R., Rosenberger, C., and Dorizzi, B. (2012b). Performance evaluation of biometric template update. In *International Biometric Performance Testing Conf.*, pages 1–4.

Jain, A., Ross, A., and Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20.

Jain, A. K., Nandakumar, K., and Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recog. Letters*, 79:80 – 105.

Kaggle (2013). Accelerometer biometric competition.

Pisani, P. H. (2017). *Biometrics in a data stream context*. PhD thesis, ICMC-USP.

Pisani, P. H. and Lorena, A. C. (2013). A systematic review on keystroke dynamics. *Journal of the Brazilian Computer Society*, 19(4):573–587.

Pisani, P. H. and Lorena, A. C. (2015). Emphasizing typing signature in keystroke dynamics using immune algorithms. *Applied Soft Computing*, 34:178 – 193.

Pisani, P. H., Lorena, A. C., and Carvalho, A. C. P. L. F. (2015). Adaptive positive selection for keystroke dynamics. *J. of Intelligent & Robotic Systems*, 80(1):277–293.

Poh, N., Kittler, J., Chan, C. H., and Pandit, M. (2015). Algorithm to estimate biometric performance change over time. *IET Biometrics*, 4(4):236–245.

Poh, N., Merati, A., and Kittler, J. (2009). Adaptive client-impostor centric score normalization: A case study in fingerprint verification. In *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–7. IEEE.

Rattani, A. (2015). *Introduction to Adaptive Biometric Systems*, pages 1–8. Springer.

Rattani, A., Marcialis, G. L., and Roli, F. (2009). *An Experimental Analysis of the Relationship between Biometric Template Update and the Doddington's Zoo: A Case Study in Face Verification*, pages 434–442. Springer Berlin Heidelberg.

Roli, F., Didaci, L., and Marcialis, G. (2008). Adaptive biometric systems that can improve with use. In Ratha, N. and Govindaraju, V., editors, *Advances in Biometrics*, pages 447–471. Springer London.

Stibor, T. and Timmis, J. (2005). Is negative selection appropriate for anomaly detection? *ACM GECCO*, pages 321–328.

Žliobaitė, I. (2010). Learning under concept drift: an overview. Technical report, Vilnius University.