IberSPEECH 2018
21-23 November 2018, Barcelona, Spain

Towards the Application of Global Quality-of-Service Metrics in Biometric Systems

J. M. Espín[1], *R. Font*[1], *J. F. Inglés-Romero*[1] *and C. Vicente-Chicote*[2]

[1]Biometric Vox S.L., Spain
[2]Universidad de Extremadura, Quercus Software Engineering Group, EPCC, Spain
{jm.espin,roberto.font,juanfran.ingles}@biometricvox.com, cristinav@unex.es

Abstract

Performance metrics, such as Equal Error Rate or Detection Cost Function, have been widely used to evaluate and compare biometric systems. However, they seem insufficient when dealing with real-world applications. First, these systems tend to include an increasing number of subsystems, e.g. aimed at spoofing detection or information management. As a result, the aggregation of new capabilities (and their interactions) makes the evaluation of the overall performance more complex. Second, performance metrics only offer a partial view of the system quality in which non-functional properties, such as user experience, efficiency or reliability, are generally ignored. In this paper, we introduce RoQME, an Integrated Technical Project funded by the EU H2020 RobMoSys project. RoQME aims at providing software engineers with methods and tools to deal with system-level non-functional properties, enabling the specification of global Quality-of-Service (QoS) metrics. Although the project is in the context of robotics software, the paper presents potential applications of RoQME to enrich the way in which performance is evaluated in biometric systems, focusing specifically on Automatic Speaker Verification (ASV) systems as a first step.

Index Terms: speaker recognition, automatic speaker verification, non-functional properties

1. Introduction

Assessing the performance of biometric systems usually involves: (1) selecting or creating an evaluation corpus, (2) defining a testing protocol, (3) applying the defined corpus and protocol to the system under consideration, and (4) computing a number of performance metrics. These metrics are generally aimed at capturing the ability of the system to produce discriminative and well-calibrated scores. The development of common corpora, protocols and metrics has been one of the major driving forces of the steady progress experienced in biometrics authentication over the last decades. This is the case, in the field of speaker recognition, of initiatives such as the ongoing *Speaker Recognition Evaluation* series [1, 2], organized by the National Institute of Standards and Technology, or the *Speakers in the Wild* [3] and the *RedDots Challenge* [4] initiatives.

The increasing complexity of biometric systems makes conventional performance metrics noticeably insufficient. Among other reasons, this is due to the interdependencies of the numerous modules that current applications involve (e.g., spoofing detection, information management, etc.). The joint operation of these modules makes it difficult to assess to what extent they contribute to the overall performance of the system. This fact has not been acknowledged until very recently [5] and there is still a lot of ground to cover in this line. In addition, and more importantly, real-world systems normally have other quality aspects often disregarded. For example, in an Automatic Speaker Verification (ASV) system, the number of failed access attempts may not only depend on the False Rejection Rate, but also on usability considerations, such as users misunderstanding the instructions and trying to authenticate with an incorrect passphrase on a text-dependent system. In the same way, the security of the system as a whole does not rest exclusively on the False Acceptance Rate, but also on some design decisions, e.g. the number of failed attempts before blocking a particular user.

Properties such as *usability* or *security* are generally referred to as *non-functional properties*. They can be considered as particular quality aspects of a system and represent an essential part of any software solution. Unfortunately, non-functional properties often get missed when designing most software systems. In this vein, the RoQME project [6], which is currently in progress, aims at enabling the modeling (at design time) and estimation (at runtime) of non-functional properties, particularly in the field of robotics. This paper introduces potential applications of the RoQME project to biometric systems, in particular, the paper focuses on ASV systems as a first attempt to address the problem. To the best of our knowledge, the evaluation of biometric systems in terms of global Quality-of-Service (QoS) metrics, defined on non-functional properties, has not been addressed before.

2. The RoQME project

The RoQME Integrated Technical Project (ITP), funded by EU H2020 RobMoSys project [7], aims at providing software engineers with a model-driven tool-chain allowing them to: (1) model system-level non-functional properties in terms of contextual information; and (2) generate ready-to-use code to evaluate QoS metrics, defined on the properties previously specified. Regarding the former, RoQME will provide engineers with a high-level language to model context variables (e.g., battery level) and, from them, relevant context patterns (e.g., "the battery level drops more than 1% per minute"). The detection of a context pattern will be considered an observation in a Bayesian Network, which is a probabilistic model used in RoQME to express the dynamics of a non-functional property (e.g., power consumption).

Once the model containing the relevant properties, contexts and observations is specified, RoQME automatically generates the code that will be responsible for estimating the degree of fulfillment of each non-funtional property in terms of QoS metrics. Basically, the implementation will consist of: (1) a context monitor that receives raw contextual data and produces context events (e.g., changes in the battery level); (2) an event processor that searches for the event patterns specified in the RoQME model and, when found, produces observations (e.g., battery is draining too fast); and, finally (3) a probabilistic reasoner that computes (based on Bayesian inference) a numeric estimation for each metric (e.g. a value of 0.89 can be understood as the probability of being optimal in terms of power consumption).

Although the RoQME project is focused on the robotics domain, the modeling tools being developed are designed to be extensible and application domain agnostic. Further details about the RoQME project can be found in [8, 9].

3. Applications to biometric systems

This section describes some potential applications of RoQME to biometric systems, in particular, to ASV systems. As stated before, although the RoQME project is focused on the robotics domain, we believe that it explores an issue of great relevance for many other software systems.

3.1. Benchmarking

Using performance metrics, such as Equal Error Rate or Detection Cost Function, to quantify the goodness of the results (either scores or binary decisions) is an essential instrument to evaluate and contrast different algorithms and technologies. However, when it comes to the integral quality of a real-world system, these metrics fail to capture some important aspects, including:

- The interaction of all the different subsystems and their combined effect on False Acceptance and False Rejection ratios.
- The effect of usability and accessibility considerations on Failure to Acquire, Failure to Enroll, and False Rejection ratios. For example, if the instructions on the screen are not clear enough, the user might be unable to provide a valid voice sample. In this case, the performance of the system would be affected before any sample reaches the verification process.
- The impact of certain system configurations and decisions at design/deployment time, such as the thresholds of the different subsystems or how many authentication attempts are granted to verify an user. Regarding the latter, it is obvious that a system with no limits on the number of failed authentication attempts would be less secure than an identical system with appropriate limits.

We propose to integrate non-functional properties (i.e., *security*, *usability*, *reliability*, etc.) into the system quality evaluation. In this sense, RoQME would allow software engineers to benchmark complex biometric systems, in terms of their overall performance, given a number of QoS metrics. Basically, engineers would start specifying the required non-functional properties by using the high-level modeling language provided by RoQME. After that, the resulting models will automatically generate the code of a software component. At runtime, this component will continuously update the QoS metrics associated with the specified properties. Finally, this information would be considered together with other performance metrics (e.g. Equal Error Rate) to compare different systems, configurations, etc.

3.2. Dynamic assessment

Monitoring QoS metrics over time would allow engineers to detect any anomaly or deviation from the expected behavior and foresee future problems. In this sense, RoQME could be useful for the early identification of malfunctioning or undesired behaviors. At runtime, RoQME can provide system administrators with real-time QoS indicators about the degree of fulfillment in usability, resource utilization or stability, to name just a few examples. This information can then be used by the engineers to improve the system or adjust its configuration.

3.3. Self-adaptation

Speaker verification systems could use the QoS metrics provided by RoQME to automatically adjust its own configuration (or software) in order to optimize its performance under changing circumstances. For instance, this approach could be applied to estimate the voiceprint quality. This would allow the system to dynamically change its operation, e.g., to ask the user to provide additional voice samples if his/her voiceprint quality falls. Furthermore, RoQME QoS metrics could play an important role in an unsupervised strategy for adaptating voiceprints, aimed at gradually improving their quality as the system is used. This approach would make the system more robust, e.g., against aging effects.

4. Acknowledgements

RoQME has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 732410, in the form of financial support to third parties of the RobMoSys project.

5. References

[1] NIST Speaker Recognition Evaluations. [Online]. Available: https://www.nist.gov/itl/iad/mig/speaker-recognition

[2] J. González-Rodríguez, "Evaluating automatic speaker recognition systems: An overview of the NIST speaker recognition evaluations (1996-2014)," *Loquens*, vol. 1, no. 1, 2014.

[3] M. McLaren, L. Ferrer, D. Castan, and A. Lawson, "The 2016 speakers in the wild speaker recognition evaluation," in *16th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2016.

[4] K. Lee, A. Larcher, G. Wang, P. Kenny, N. Brümmer, D. A. van Leeuwen, H. Aronowitz, M. Kockmann, C. Vaquero, B. Ma, H. Li, T. Stafylakis, M. J. Alam, A. Swart, and J. Perez, "The reddots data collection for speaker recognition," in *15th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2015.

[5] T. Kinnunen, K.-A. Lee, H. Delgado, N. W. D. Evans, M. Todisco, M. Sahidullah, J. Yamagishi, and D. A. Reynolds, "t-DCF: a detection cost function for the tandem assessment of spoofing countermeasures and automatic speaker verification," in *ODYSSEY 2018 – The Speaker and Language Recognition Workshop*, 2018.

[6] RoQME Integrated Technical Project. (2018-2019). RoQME: Dealing with non-functional properties through global Robot Quality-of-Service Metrics. [Online]. Available: http://robmosys.eu/roqme/

[7] RobMoSys EU H2020 Project. (2017-2020). RobMoSys: Composable Models and Software for Robtics Systems - Towards an EU Digital Industrial Platform for Robotics. [Online]. Available: http://robmosys.eu

[8] C. Vicente-Chicote, J. F. Inglés-Romero, and J. Martínez, "A component-based and model-driven approach to deal with non-functional properties through global qos metrics," in *5th International Workshop on Interplay of Model-Driven and Component-Based Software Engineering (ModComp), in conjunction with MODELS'18*, 2018.

[9] J. F. Inglés-Romero, J. M. Espín, R. Jiménez-Andreu, R. Font, and C. Vicente-Chicote, "Towards the use of quality of service metrics in reinforcement learning: A robotics example," in *5th International Workshop on Model-driven Robot Software Engineering (MORSE), in conjunction with MODELS'18*, 2018.