JCC-BD & ET 2022

Short Papers

of the

10th Conference on Cloud Computing, Big Data & Emerging Topics

HTTPS://JCC.INFO.UNLP.EDU.AR

@CONF_CC_BD_ET

JCC@LIDI.INFO.UNLP.EDU.AR

POSTGRADO
FACULTAD DE INFORMÁTICA

III-LIDI
INSTITUTO DE INVESTIGACIÓN EN INFORMÁTICA - LIDI

Short Papers of the 10th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2022)

Marcelo Naiouf
Armando De Giusti
Franco Chichizola
Laura De Giusti
Enzo Rucci
- Editores -

La Plata, Buenos Aires, Argentina.
June 28−30, 2022

Short Papers of the 10th Conference on Cloud Computing, Big Data & Emerging Topics / editado por Armando De Giusti ... [et al.]. - 1a ed. - La Plata: Universidad Nacional de La Plata. Facultad de Informática, 2022.
Libro digital, PDF/A

Archivo Digital: descarga y
online ISBN 978-950-34-2126-0

ISBN 978-950-34-2126-0

1. Computación. 2. Actas de Congreso. 3. Conferencias. I. De Giusti, Armando, ed.
CDD 004.071

Preface

Welcome to the short paper proceedings of the 10th Conference on Cloud Computing, Big Data & Emerging Topics (JCC-BD&ET 2022), held in a hybrid modality (both on-site and live online settings were allowed). JCC-BD&ET 2022 was organized by the III-LIDI and the Postgraduate Office, both from School of Computer Science of the National University of La Plata.

Since 2013, this event has been an annual meeting where ideas, projects, scientific results and applications in the cloud computing, big data and other related areas are exchanged and disseminated. The conference focuses on the topics that allow interaction between academia, industry, and other interested parties.

JCC-BD&ET 2022 covered the following topics: cloud, edge, fog, accelerator, green, web and mobile computing; big and open data; machine and deep learning; smart and sustainable cities, and special topics related to emerging technologies. In addition, special activities were also carried out, including 1 plenary lecture and 1 discussion panel.

In this edition, 16 short papers were accepted after the peer-review process. These short papers correspond to initial research with preliminary results, ongoing R+D projects, or postgraduate thesis proposals. The authors of these submissions came from the following 5 countries: Argentina, Chile, Ecuador, Panama, and Spain. We hope readers will find these contributions useful and inspiring for their future research.

Special thanks to all the people who contributed to the conference's success: program and organizing committees, authors, reviewers, speakers, and all conference attendees.

June 2022

Marcelo Naiouf
Armando De Giusti
Franco Chichizola
Laura De Giusti
Enzo Rucci

Organization

General Chair

Marcelo Naiouf	Universidad Nacional de La Plata, Argentina.

Associate Chairs

Armando De Giusti	Universidad Nacional de La Plata and CONICET, Argentina.
Franco Chichizola	Universidad Nacional de La Plata, Argentina.
Laura De Giusti	Universidad Nacional de La Plata and CIC, Argentina.
Enzo Rucci	Universidad Nacional de La Plata and CIC, Argentina.

Program Committee

María José Abásolo	Universidad Nacional de La Plata and CIC, Argentina
José Aguilar	Universidad de Los Andes, Venezuela
Jorge Ardenghi	Universidad Nacional del Sur, Argentina
Javier Balladini	Universidad Nacional del Comahue, Argentina
Oscar Bria	Universidad Nacional de La Plata and INVAP, Argentina
Silvia Castro	Universidad Nacional del Sur, Argentina
Laura De Giusti	Universidad Nacional de La Plata and CIC, Argentina
Mónica Denham	Universidad Nacional de Río Negro and CONICET, Argentina
Javier Diaz	Universidad Nacional de La Plata, Argentina
Ramón Doallo	Universidade da Coruña, Spain
Marcelo Errecalde	Universidad Nacional de San Luis, Argentina
Elsa Estevez	Universidad Nacional del Sur and CONICET, Argentina
Aurelio Fernandez Bariviera	Universitat Rovira i Virgili, Spain
Héctor Florez Fernández	Universidad Distrital Francisco José de Caldas, Colombia
Fernando Emmanuel Frati	Universidad Nacional de Chilecito, Argentina

Carlos Garcia Garino	Universidad Nacional de Cuyo, Argentina
Carlos García Sánchez	Universidad Complutense de Madrid, Spain
Adriana Angélica Gaudiani	Universidad Nacional de General Sarmiento, Argentina
Graciela Verónica Gil Costa	Universidad Nacional de San Luis and CONICET, Argentina
Roberto Guerrero	Universidad Nacional de San Luis, Argentina
Waldo Hasperué	Universidad Nacional de La Plata and CIC, Argentina
Francisco Daniel Igual Peña	Universidad Complutense de Madrid, Spain
Tomasz Janowski	Gdansk University of Technolgy, Poland
Laura Lanzarini	Universidad Nacional de La Plata, Argentina
Guillermo Leguizamón	Universidad Nacional de San Luis, Argentina
Edimara Luciano	Pontificia Universidade Católica do Rio Grande do Sul, Brazil
Emilio Luque Fadón	Universidad Autónoma de Barcelona, Spain
Mauricio Marín	Universidad de Santiago de Chile, Chile
Luis Marrone	Universidad Nacional de La Plata, Argentina
Katzalin Olcoz Herrero	Universidad Complutense de Madrid, Spain
José Angel Olivas Varela	Universidad de Castilla-La Mancha, Spain
Xoan Pardo	Universidade da Coruña, Spain
Patricia Pesado	Universidad Nacional de La Plata, Argentina
María Fabiana Piccoli	Universidad Nacional de San Luis, Argentina
Luis Piñuel	Universidad Complutense de Madrid, Spain
Adrian Pousa	Universidad Nacional de La Plata, Argentina
Marcela Printista	Universidad Nacional de San Luis, Argentina
Dolores Isabel Rexachs del Rosario	Universidad Autónoma de Barcelona, Spain
Nelson Rodríguez	Universidad Nacional de San Juan, Argentina
Juan Carlos Saez Alcaide	Universidad Complutense de Madrid, Spain
Aurora Sánchez	Universidad Católica del Norte, Chile
Victoria Sanz	Universidad Nacional de La Plata, Argentina
Remo Suppi	Universidad Autónoma de Barcelona, Spain
Francisco Tirado Fernández	Universidad Complutense de Madrid, Spain
Juan Touriño Dominguez	Universidade da Coruña, Spain
Viale Pereira, Gabriela	Danube University Krems, Austria
Gonzalo Zarza	Globant, Argentina

Sponsors

Sistema Nacional
de Computación
de Alto
Desempeño

Table of Contents

Fog, Edge and High-Performance Computing

IoT for agriculture optimization: preliminary results of a tropical precision farming project

Soizic Gibeaux[1,2][0000-0002-5833-3604], Javier Pitti[3] and Carlos Gonzalez[1,2][0000-0002-4892-4451]

[1] Vicerrectoría Investigación y Posgrado, Universidad Autónoma de Chiriquí, David, Panamá.
[2] Sistema Nacional de Investigación, SENACYT, Ciudad del Saber, Panamá.
[3] Instituto de Innovación Agropecuaria de Panamá IDIAP, Cerro Punta, Panamá.
alice.soizic_1@unachi.ac.pa, javier.pitti@idiap.gob.pa,
carlos.gonzalez5@unachi.ac.pa

Abstract. The agricultural activities need optimal soil and the agro-climatic conditions in the production areas such as Cerro Punta in Panama. Among the innovative technological alternatives, there are tools that can provide farmers with information describing the agro-meteorological conditions of some areas, such as precision agriculture. Due to the data range of wireless communication and low power capabilities, one of the most promising technologies for precision farming is WSN with integrated LoRa systems and energy self-sufficient. This study presents the methodology of the first LoRa WSN project destined to precision agriculture in Panama, and bring the preliminary results in the evaluation of LoRa signal, micro solar panel efficiency and environmental data. In the tropical vegetation of this part of the country, the range of LoRa communication is smaller than theoretically expected. The efficiency of the solar panels is enough to maintain the batteries always charged. Finally, slightly different microclimates have been highlighted between two close monitored areas.

Keywords: Environmental monitoring, Internet of Things, Wireless Sensors Network, LoRaWAN, Received Signal Strength Indicator, Precision farming.

1 Introduction

The development of agricultural activities is closely linked to the state of the soil and the agro-climatic conditions of the production areas. Specific actions such as the estimation of river precipitation, temperature, photoperiod, physicochemical properties of the soil, among other variables, make it possible to identify the potential and limitations of different areas in a timely manner.

Precision agriculture is a valuable tool for accurately diagnosing agricultural production problems, making decisions and obtaining satisfactory responses in agricultural yield indices. It is vitally important to focus on the right place at the right time, based on the scientific innovations offered by information technology.

In the IoT environment, the technological development of Wireless Sensor Networks (WSN) offers a sustainable solution applied to precision agriculture. This type of

network allows efficient use of agricultural resources, management tools and monitoring of different parameters to achieve higher quality yields and production [1]. The concept WSN can be extended under low-power wide area networks, a wireless technology for transmitting small data units over long distances with minimal power consumption. In a precision agriculture environment, the scope of LoRaWAN can be used to monitor crop health forecasting, ensuring suitable amounts of nutrients, disease detection, irrigation scheduling and weather monitoring [2].

Some research contributions present technical and theoretical approaches, mainly focused on the data collection, data analysis, diagnosis of forecast disease and field operation of precision agriculture techniques [3-5].

The development of institutional cooperation alliances is fundamental for the fulfillment of strategic objectives in the field of agricultural research and innovation. The Institute for Agricultural Innovation of Panama (IDIAP), in its mission to strengthen the national agro-technological base, together with the Autonomous University of Chiriqui (UNACHI), has achieved the procurement of sensors through the FIED19-R1-003 project, with the interest of linking Internet of Things (IoT) technologies to the experimental trials that are conducted daily at the IDIAP Experimental Station, located in Cerro Punta, District of Tierras Altas, Province of Chiriqui.

2 Materials and Methods

The experiments were managed in an experimental station of the Institute for Agricultural Innovation of Panama (IDIAP), located in Cerro Punta. The design of the IoT framework requires sensor nodes, LoRaWAN gateway, LoRaWAN Network Server and a monitoring solution for IoT applications. The experiments were conducted over onions parcels which combines practical WSNs deployment, with precise applicable solutions on the crops.

2.1 Wireless Sensors Network

Table 1. Parameters of LoRa nodes.

Parameters	Values
Spreading factor	7-12
Coding rate	4/5
Bandwidth	125 kHz
Frequency range	902-928 MHZ
Transmission power	-2.0 dBm

The implementation of LoRaWAN was chosen due to its low data transfer rate, low deployment and management costs. In tropical areas, agricultural plots are susceptible to different types of interference and diversification, requiring an improved LoRaWAN network design scheme. The IoT nodes devices are divided in tree crop areas, with a

size range coverage of 1,000 m^2 and the gateway has a height of 3 m. The packets collected are transmitted through a LoRa network communication and then forwarded to LoRaWAN Gateway with LTE network connection. In order to transmit the packets, the LoRa Nodes adopt the spreading factor automatically between SF 7-12. The main parameters in field trials are given in **Table 1**.

2.2 Cerro Punta site case study

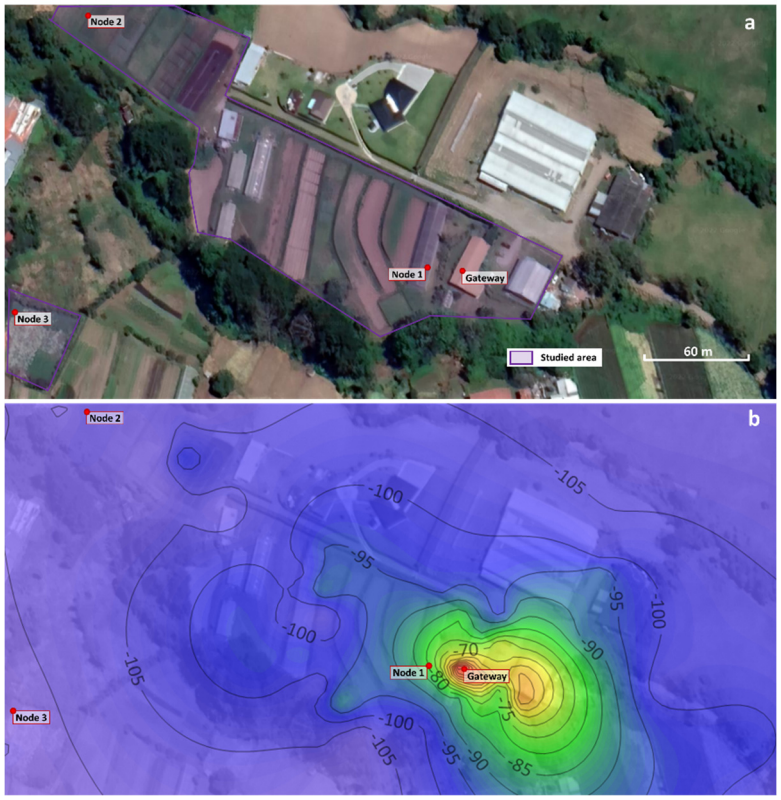


Figure 1. Sensor nodes and Gateway locations (a) and RSSI signal estimation map (b).

Cerro Punta is an important productive area of the country of Panama, the installed sensors have the capacity to collect data on parameters directly related to the productive cycle, development and growth of the main crops evaluated in the Highlands area.

In crops such as onions, the development of diseases such as Botrytis is favored by conditions of periods of high relative humidity and moderate temperatures. Periods of about six hours of wet leaves and temperatures below 24°C are sufficient for infection to occur. Under these conditions the fungal growth organisms (spores) have favorable conditions to reproduce [6]. Evapotranspiration is also affected by climate, management and crop development environment. The main climatic parameters affecting evapotranspiration are solar radiation, temperature and relative air humidity [7].

The Node 1 was placed close to the Gateway, at 20m, in a greenhouse of 4m high, 45m long and 9m wide while Nodes 2 and 3 were placed at 250m from the gateway, and 180m from each other in outdoor 400m^2 parcels (Fig.Figure **1**.a).

3 Preliminary results

To quantify the signal strength, an estimated isopleth has been created using Surfer software with Natural Neighbor as gridding method to simulate a signal strength map (Fig.Figure **1**.b). Comparing the RSSI results with those of a similar study with a max range around 200m against 400m for our study, Yim *et al.* [8] obtained a RSSI of -92, whereas we get a RSSI of -110 without significant packet-loss.

As for the efficiency of the solar panels in recharging the batteries, it was noted that with an average luminosity value of 8,000 lux.h^{-1}, the average charge rate is 0.03 V.h^{-1}, which represents 1% per hour of the maximum charge of the battery during 12 hours. The average discharge of the batteries is 0.01 V.h^{-1} without luminosity, during 12 hours of the night. As a result, the night discharge is completely compensated during the day.

In the greenhouse, node 1 detected low light and very low soil moisture compared to the outdoor sensors. In the latter two situations, nodes 2 and 3, demonstrated more diverse air temperature values than expected. Indeed, despite the short distance between the two nodes, it has been detected different variations over time between the two monitored sites, especially in the extreme temperature values in which node 3 showed a minimum of 1.17°C lower than node 2, and a maximum 4.60°C higher while relative humidity, soil moisture and soil temperature variations are the same on both sites. Sensor parameters and obtained values are detailed in Table Table **2**.

Table 2. Sensor nodes measurements during one week from February 25 to March 4, 2022.

Parameters		**Node 1**	**Node 2**	**Node 3**
RSSI signal (dB)	Average ± Std Dev	-80 ± 1	-113 ± 2	-112 ± 3
Battery (V)	Average ± Std Dev	4.02 ± 0.10	4.03 ± 0.08	4.00 ± 0.09
	Min - Max	3.81 - 4.20	3.90 - 4.19	3.86 - 4.18
Luminosity (Lux)	Average ± Std Dev	5,237 ± 9,946	7,606 ± 14,647	8,845 ± 15,526
	Min - Max	0 – 54,612	0 – 54,612	0 – 54,612
Temperature (°C)	Average ± Std Dev	19.13 ± 10.16	16.35 ± 7.31	15.44 ± 8.44
	Min - Max	5.03 - 46.80	4.39 - 35.90	3.22 – 40.50
Soil temperature (°C)	Average ± Std Dev	23.14 ± 5.76	18.51 ± 4.96	18.95 ± 5.55
	Min - Max	14.30 – 39.10	9.93 – 31.00	9.36 – 33.90
Relative Humidity (%)	Average ± Std Dev	74 ± 29	81 ± 21	81 ± 24
	Min - Max	14 - 100	26 - 100	19 – 100
Soil moisture (%)	Average ± Std Dev	5.6 ± 2.8	38.6 ± 11.3	40.13 ± 11.4
	Min - Max	0.1 – 11.0	30.5 – 74.2	30.4 – 74.3

4 Conclusion and future work

The aim of this paper was to assess the functioning of the first LoRa WSN destined to precision agriculture and farmers decision-making assistance in Panama. The preliminary results have been analyzed during the first week of monitoring. The signal range was smaller than in other studies due to the high vegetation density of the area. In terms on energy, the tropical environment is well adapted to the use of solar panels and the night discharge of the batteries was totally compensated during the day. From an environmental point of view, the microclimates differences have already been highlighted between two close monitored areas at less than 180 meters from each other. It has been observed up to 5°C differences in the air while soil temperature, soil moisture and relative humidity maintained stable between the two sites. Nonetheless, those differences suggested the necessity to adapt the local management of these two parcels. As future work, to obtain a better coverage range, LoRaWAN antennas with ranges of up to 20 km can be implemented and the acquisition of more data on local environmental conditions will allow the analysis of plant interactions contributing to the management of diseases, water use and the use of phytosanitary products and fertilizers.

Aknowledgements

This work was made possible thanks to the support of the Sistema Nacional de Investigación and (FIED19-R1-003) of Secretaría Nacional de Ciencia, Tecnología e Innovación Senacyt Panamá.

References

1. Ofori, M., El-Gayar, O.: Drivers and challenges of precision agriculture: a social media pespective. Precision Agriculture 22 (3), 1019–1044 (2021).
2. Davcev, D., Mitreski, K., Trajkovic, S., Nikolovski, V., Koteli, N.: IoT agriculture system based on LoRaWAN. In: 2018 14th IEEE International Workshop on Factory Communication Systems (WFCS), pp. 1-4. (2018).
3. Reka, S. S., Chezian, B. K., Chandra, S. S.: A novel approach of iot-based smart greenhouse farming system. Green buildings and sustainable engineering, 227–235 (2019).
4. Delnevo, G., Girau, R., Ceccarini, C., Prandi, C.: A deep learning and social iot approach for plants disease prediction toward a sustainable agriculture. IEEE Internet Things Journal, (2021).
5. Yang, J., Sharma, A., Kumar, R.: IoT-based framework for smart agriculture. International Journal of Agricultural and Environmental Information Systems IJAEIS 12(2), 1–14 (2021).
6. Marcuzzo, L. L., Haveroth, R.: Development of a weather-based model for Botrytis leaf blight of onion. Summa phytopathologica 42, 92-93 (2016).
7. Wang, Z., et al.: Spatiotemporal variability of reference evapotranspiration and contributing climatic factors in China during 1961–2013. Journal of Hydrology 544, 97-108 (2017).
8. Yim, D., et al.: An experimental LoRa performance evaluation in tree farm. In: 2018 IEEE sensors applications Symposium (SAS), pp. 1-6. (2018, March).

Incorporating Resilience to Platforms based on Edge and Fog Computing

Santiago Medina [(1)], Diego Montezanti [(1)], Lucas Gómez D´Orazio[(1)], Evaristo Compagnucci[(1)], Armando De Giusti [(1)(2)], Marcelo Naiouf [(1)]

[1] Instituto de Investigación en Informática LIDI (III-LIDI),
Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
[2] CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas

{smedina,dmontezanti,degiusti,mnaiouf}@lidi.info.unlp.edu.ar
{evaristogaston,lucas.dorazio}@alu.ing.unlp.edu.ar

Abstract. In recent years, Internet of Things (IoT) has become extremely popular because of its ability of sensing information from the environment and processing it in the cloud. Edge and Fog Computing are new paradigms that aim to localize some of the processing near the sensors, helping to cope with high communications latencies and bandwidth bottlenecks. As Wireless Sensor Networks (WSN) and Fog nodes are prone to failure, affecting system reliability and performance, the implementation of resilience strategies becomes essential to ensure reliable delivery of data and system availability during interruptions or in the presence of faults. In this article, we present three lines of research in progress: Redundant Image Processing, Integration of a WSN with an IoT Platform for Intelligent Control and Resilient Monitorization and Control of Robots. We aim to incorporate resilience mechanisms to platforms that integrate Edge, Fog and Cloud Computing, and to evaluate the proposed solutions in terms of the coverage achieved, processing and communication times and power consumption.

Keywords: Edge Computing, Fog Computing, Cloud Computing, Resilience, Internet of Things, Wireless Sensor Networks.

1 Introduction/Motivation

In recent years, the Internet of Things (IoT) has gained a lot of attention because of its characteristic of typically enabling the connection of a significant number of sensor devices that sense information from the environment and share it to a cloud service for processing [1]. This has been extensively used to develop smart applications, such as traffic management, smart houses, monitoring of natural events and human health [2]. Due to the growing popularity of the IoT, the number of Internet-connected devices has increased significantly. As a result, a huge amount of network traffic is generated, which may lead to bottlenecks, and eventually generate limitations in terms of

communication latency with the cloud and network bandwidth [3], so the traditional cloud-based infrastructures are not enough for the current demands of IoT applications [4]. To deal with these issues, in recent years, the paradigms of Fog computing and Edge computing were proposed to alleviate these limitations, by moving some processing capabilities closer to the network edge and away from the central cloud servers [1]. This allows to distribute the computations of the IoT data, and to reduce the communication latency [3].

In IoT systems, data are acquired by wireless sensor networks (WSNs), which are deployed in harsh environments where weather and other factors can cause node failure. In addition, the IoT devices and the nodes in WSNs are heterogeneous, highly distributed, reliant on wireless communications and generally powered by batteries, which are prone to failure [2]. All of this makes the recovery of devices, and the creation of a pattern for Fault Tolerance in IoT, especially difficult [4].

In the Fog layer, some of the computing nodes may be unreliable and fail unexpectedly, affecting the system's reliability. For this reason, mechanisms for handling node failures become essential but are also especially challenging, because when a Fog node fails, moving the computations to neighbor nodes (or to the cloud) may increase the communication latency, and thus affect the system performance [3]. Consequently, the design of an effective fault tolerance mechanism is crucial to ensure reliable delivery of data and to warrant the system availability during interruptions of any kind, or in the presence of faults [2].

Resilience can be introduced at different architectural layers because a fault can occur at any of the layers. Either sensors and actuators, network, or computation and storage nodes, can perform erroneously in their layers [5].

2 Main goals

Our main goal in this project is to incorporate mechanisms aimed at obtaining a certain degree of resilience on a platform that integrates sensor networks with levels of Edge, Fog and Cloud Computing (see Fig. 1).

To do this, we propose to evaluate different scenarios of communication failures on such a 3-tier architecture. The end nodes are sensors connected to microcontrollers, robots, drones, or other devices that can perform tasks remotely controlled by a server.

Faced with possible communication failures between the different levels of the architecture, it is planned to redundantly process information in each of the layers, so that the system can maintain a degree of responsiveness. Through the measurement of response times and energy consumption, we hope to obtain criteria to evaluate the convenience of carrying out the processing at a specific level, but also of having the alternative of transferring it to another layer in presence of a communication failure, in order to provide fault coverage.

In the case of communication failures between robots (or end devices) and servers, validating the status is proposed. This validation can be either periodical or event-driven (like the fulfillment of partial objectives or the reaching of intermediate posi-

tions), by exchanging round trip messages with the server. Therefore, the device could return to the last consistent state and wait for the link to restore. For this, the valid states must be stored in the final device, so that the recovery can be autonomously made.

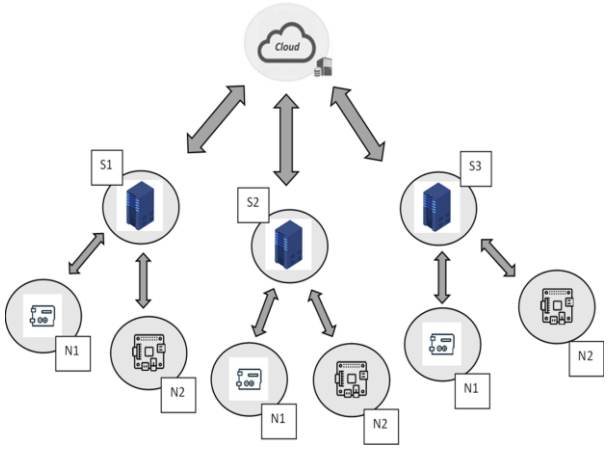


Fig. 1. 3-layer architecture

Accordingly, in both cases, we aim to improve the integrity of the system, allowing it to maintain some basic functionality in the presence of communication failures, or at least to remain in a safe state, by adopting the most appropriate strategy for each case, similar to that explored in [6]. The deliberate provocation of communication failures will allow us to evaluate the robustness of the proposed solutions.

3 Works in progress

3.1 Redundant Image Processing in 3-layer architecture

In this line of research, the focus is on the conceptual integration of resilience aspects in an architecture with Edge, Fog and Cloud Computing levels.

In the 3-layer platform that we propose, the Edge level consists of a node built from an ESP32-CAM [7], which takes pictures when movement is detected. The pictures are transmitted to a Fog server, which can pre-process them and, in turn, transmit them to the Cloud to be able to take a concrete action from the Internet. As each of the three levels has a computational capacity, certain data processing tasks can be performed alternatively in any of them. This makes it possible to implement resilience strategies based on redundancy.

There are two possible scenarios: in normal operation, our goal is to evaluate processing times and communication latencies, in order to decide where to process the

images to detect people or objects. However, in the presence of communication failures, we aim to take advantage of the computing power, whether in the Fog server or in the sensor node itself, to redundantly process the images to maintain functionality. As a consequence, we will be able to characterize the system performance in terms of processing times at each level, coverage obtained against failures and energy consumption.

3.2 Integration of a WSN with an IoT Platform for Intelligent Control of Classrooms

In this line of research, we are studying the deployment and configuration of the IoT platform ThingsBoard [8], to be integrated with a Wireless Sensor Network (WSN) built with CO2 sensors and energy consumption measurement and control nodes, in the context of a university building with classrooms that are monitored and remotely controlled [9].

The CO2 sensors in each classroom are connected via WiFi to a Raspberry Pi (which is at the Fog level) that reads the data (via HTTP), and makes some preprocessing before transmitting them to the ThingsBoard server. This Fog server can maintain centralized monitorization of the CO2 levels across the whole building, and activate alerts in case of undesired values.

The energy consumption measurement and control nodes in each classroom are directly connected via WiFi to the Fog ThingsBoard server, publishing information in MQTT [10] topics. The server can maintain centralized monitorization of power consumption and take concrete actions, such as remotely turning off lights or air conditioning equipment when the classroom is not being used.

3.3 Resilient Monitorization and Control of Robots

Our goal in this line of research is to incorporate resilience to a system that uses Lego Mindstorms EV3 robots [11] to perform specific tasks controlled by a server. We aim to develop an application that controls the functions of the robots through resilient communications.

Currently, we are exploring 3 alternative strategies, in the attempt to maintain the consistency of the system in the presence of failures. Each of them has particular characteristics as regards to detection latency, coverage and workload to be rerun.

1. The server communicates the whole information of the task to be performed or the path to be followed. The robot must store the path and follow it autonomously, communicating with the server upon completion. If the connection is lost during the trajectory, the task cannot be validated. The only resilience strategy is to restart from the beginning if the final point does not match with the expected one. This variant minimizes communication load but maximizes the detection latency.
2. The server communicates partial information of the task or the path. The robot must store the partial path and communicate with the server when reaching the target point, which in turn validates the completed chunk. If the connection is lost, the robot has to use the stored information to return to the starting point, which is the

last valid one. Compared with the previous variant, this one involves more communications but improves the detection latency.

3. The server communicates partial information of the task or the path. In addition, a periodic keepalive signal is transmitted via an independent socket or a different MQTT topic. The robot must store the partial path and communicate with the server when reaching the target point, which in turn validates the completed chunk. In the meanwhile, the status of the connection is monitored with periodical messages. Consequently, if the connection is lost, the robot has to use the stored information to return to the last valid point. This latter alternative involves frequent messages but minimizes the detection latency.

4 Conclusions

Although the research is in an incipient stage, this work proposes to implement mechanisms to provide robustness to a system against communication failures, and to determine the performance of the different resilience strategies in terms of the coverage achieved, processing and communication times and power consumption.

References

1. Mohan, N., & Kangasharju, J. (2016, November). Edge-fog cloud: A distributed cloud for internet of things computations. In 2016 Cloudification of the Internet of Things (CIoT) (pp. 1-6). IEEE.
2. Tong, Y., Tian, L., Lin, L., & Wang, Z. (2020). Fault Tolerance Mechanism Combining Static Backup and Dynamic Timing Monitoring for Cluster Heads. IEEE Access, 8, 43277-43288.
3. Karagiannis, V., Desai, N., Schulte, S., & Punnekkat, S. (2020). Addressing the node discovery problem in fog computing. In 2nd Workshop on Fog Computing and the IoT (Fog-IoT 2020). Schloss Dagstuhl-Leibniz-Zentrum für Informatik.
4. Bierzynski, K., Escobar, A., & Eberl, M. (2017, May). Cloud, fog and edge: Cooperation for the future?. In 2017 Second International Conference on Fog and Mobile Edge Computing (FMEC) (pp. 62-67). IEEE.
5. Agrawal, A., & Toshniwal, D. (2021). Fault Tolerance in IoT: Techniques and Comparative Study. Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146, 7(1), 49-52.
6. Montezanti, D., Rucci, E., De Giusti, A., Naiouf, M., Rexachs, D., & Luque, E. (2020). Soft errors detection and automatic recovery based on replication combined with different levels of checkpointing. Future Generation Computer Systems, 113, 240-25.
7. Espressif Web Page, https://www.espressif.com/en/news/ESP32_CAM, last accessed 2022/04/28.
8. ThingsBoard Homepage, https://thingsboard.io, last accessed 2022/04/27.
9. De Antueno, J., Medina, S., De Giusti, L., & De Giusti, A. (2020). Analysis, Deployment and Integration of Platforms for Fog Computing. Journal of Computer Science and Technology, 20(2), e12-e12.
10. MQTT Homepage, https://mqtt.org/, last accessed 2022/04/15.
11. ev3dev Homepage, https://www.ev3dev.org/, last accessed 2022/04/08.

Comparative analysis of exhaustive searching on a massive finger-vein database over multi-node/multi-core and multi-GPU platforms

Sebastián Guidet[1], Ruber Hernández-García[2], Fernando Emmanuel Frati[1], and Ricardo J. Barrientos[2,3]

[1] Department of Basic and Technological Sciences,
Universidad Nacional de Chilecito, Chilecito, La Rioja, Argentina
sguidet@undec.edu.ar, fefrati@undec.edu.ar
[2] Laboratory of Technological Research in Pattern Recognition (LITRP),
Faculty of Engineering Sciences, Universidad Católica del Maule, Talca, Chile
rbarrientos@ucm.cl, rhernandez@ucm.cl
[3] Department of Computer Science and Industries, Faculty of Engineering Science,
Universidad Católica del Maule, Talca, Chile

Abstract. When searching on unstructured data (video, images, etc.), response times are a critical factor. In this work we propose an implementation on two types of multi-GPU and multi-node/multi-core platforms, for massive searches. The presented method aims to reduce the time involved in the search process by solving simultaneous queries over the system and a database of millions of elements. The results show that the multi-GPU approach is 1.6 times superior to the multi-node/multi-core algorithm. Moreover, in both algorithms the speedup is directly proportional to the number of nodes reaching 156x for 4 GPUs, and 87x in the case of the hybrid multi-node/multi-core algorithm.

Keywords: High Performance Computing, identification of individuals, Local linear binary pattern, Finger veins, GPU.

1 Introduction

The volume of data in the world is growing exponentially. According to some estimates by the united nations, 90% of the world's data has been created in the last two years and is predicted to grow at 40% per year. In 2020, 64.2 zettabytes of data were created, a 314% increase over 2015 [7]. Most of this data is unstructured (videos, images, etc.), so by needing to perform a search it cannot be treated in the same way as traditional databases. In multimedia databases it is not possible to perform exact searches, because the information is not always stored in the same way. Two images may look the same, but when compared pixel by pixel they are totally different. In this type of databases, similarity search is used, which consists of retrieving all those objects within a database that are similar according to a given query.

Similarity searches are mathematically modeled through metric spaces using a distance function [1]. The distance function is responsible for determining the

degree of similarity between two objects. As the size of the database increases, the system must perform more distance calculations. Consequently, the performance of similarity searches is largely conditioned by the distance function computation and the size of the database, which influences the total processing time.

Due to the computational load involved in the search process and its impact on the processing time, it is necessary to search for alternatives to speed up the computation involved in solving queries on the multimedia database. One of the most widely used methods in the literature to reduce costs in terms of time is through parallel processing. Although the parallelization of algorithms is not a new topic, the emergence of GPU coprocessors nowadays allows a high level of parallelism at a very low cost.

In this paper, we propose a comparison between multi-node/multi-core and multi-GPU parallel methods, applying an exhaustive search algorithm to solve k-NN queries in metric spaces. As a case study we use a database of finger vein images composed of 40,000,000 images, preprocessed by the CLAHE (limited contrast histogram adaptive histogram equalization) algorithm and the vertical LLBP (vertical linear local binary pattern) descriptor. In the experiments performed we used the Hamming distance to measure the similarity between two images.

2 Background on similarity search

The search for objects in a database that are similar to a given query object is a problem that has been extensively studied in recent years. The solutions are based on two criteria: the first one determines how the selection of the solution set will be performed when searching; the second one refers to the way in which the database is traversed to apply the first criterion.

For the selection of the solution set, the model of metric spaces [1] is used. A metric space (X, d) consists of a universe of valid objects X and a distance function $d : XxX \rightarrow R^+$ defined between them. The distance function determines the similarity between two given objects. The finite subset $U \subset X$ with size $n = |U|$, is called the database and represents the collection of objects in the search space. There are two main queries of interest, kNN and range queries.

Range query[1]: the objective is to retrieve all objects $u \in U$ within a radius r of the query q.

The k nearest neighbors (kNNN)[2]: the goal is to recover the set $kNNN(q) \subseteq U$ such that $|kNNN(q)| = k$ and $forallu \in kNN(q); v \in U - kNN(q); d(q, u) \leq d(q, v)$.

When working with large databases, even if the radius is small, range queries provide large solution sets [4]. Therefore, in these cases the most commonly used and efficient query method is k-nearest neighbor queries. However, when applying the k-nearest neighbors method, it is necessary to calculate the distance between the query and each element belonging to the database. For this calculation, the Hamming distance similarity function is used, whose effectiveness has been proven by other authors in the literature [8].

When performing the database traversal to resolve similarity queries in metric spaces, the most trivial but costly (in time and/or resources) method is the brute

force method. An alternative is the index-based methods [3], but they have the disadvantage of losing efficiency as the size of the database increases, in addition to presenting problems in shared memory systems [2].

3 Searching process on a massive database

The identification of individuals consists of an exhaustive 1:N search in the database, this procedure returns a list of 32 records sorted by similarity score in ascending order. We only get the first 32 results because it is the lowest perfect recognition rank for $LLBPv$ with the best [6] precision performance.

This process must be performed for each query received by the system. Thus, the system workload increases with a high rate of queries per unit time, and the volume of data to be processed increases significantly, so the response time must be reduced.

3.1 Multi-node/multi-core searching algorithm

The presented approach attempts to reduce the computation time of 1:N similarity comparisons by achieving a significant speedup for an exhaustive search process on a massive database. For this, a hybrid multi-node/multi-core parallel version was implemented in order to distribute the tasks among the different nodes of the cluster. Therefore, each node computes the similarity tests on a partition of the database. Each node handles its tasks by applying the Round Robin ($dist - round - Robins$) distribution scheme [6]. Each query is solved with 8 threads, as each node has 36 available threads, each node of the algorithm can process up to 4 queries in parallel.

Once all the processes have completed their tasks in parallel, a single array is generated with the 32 elements of smallest distances in increasing order. Finally, each node sends its 32 shortest distance results (local results) to the master node, which computes the final 32 results.

3.2 Multi-GPU searching algorithm

In order to process a larger amount of items in the shortest possible time, a multi-GPU algorithm is implemented, where a multi-threaded session is started on the CPU to handle each GPU (one GPU per thread). The algorithm divides the database into parts, each part will be processed by a different GPU until the number of items in the database is completed. It should be noted that each GPU processes its portion of the database by implementing the algorithm described in [5], where each query is solved with a different CUDA block. Finally, the quicksort algorithm is used to sort and return the 32 smallest distances of the process. It should be noted that each GPU of the multi-GPU platform can only process 56 simultaneous queries due to the occupancy factor of the model used [5].

4 Experimental results

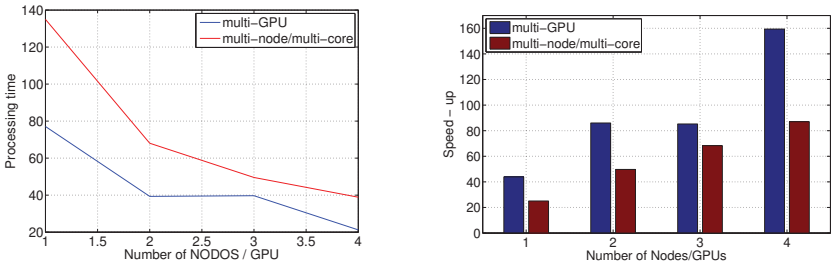
The experimental environment is composed of 4 servers (or nodes), and each node with 2 Intel Xeon Gold 6140 CPUs @ 2.30 GHz, totaling 36 physical cores, 24.75 MB of L3 cache and 126 GB of RAM. In addition, one of these servers

has 4 GPU model NVIDIA GeForce GTX 1080 TI, CUDA Cores: 3584, GPU Memory: GDDR5X 11GB.

To evaluate the ability of the proposed algorithm to respond to simultaneous queries in adequate time, the BigFVDB dataset generated in previous work is used [6], using 4,000,000 samples for these experiments.

The experiments were performed by increasing the number of processing nodes in both architectures up to 4 nodes. To obtain an unbiased result and ensure the stability of the results, the time measurements were averaged by repeating each test 100 times. In addition, it was verified that in all experiments the same results were obtained for the same comparisons.

Figure 1 (a) summarizes the results obtained. It should be noted that for the experiments the number of simultaneous queries was set to 56 in order not to exceed the occupancy factor of the GPUs used. As can be seen in both cases the processing times decrease as the number of nodes in the system increases, with the multi-GPU algorithm being 1.6 times faster than the multi-node/multi-core algorithm.



(a) Processing time (seconds). (b) Speed-up

Fig. 1: Processing time and speed-up when solving kNN queries (k=32) for the multi-GPU algorithm versus the multi-node/multi-core algorithm for 56 simultaneous queries and 4,000,000 items. The X-axis shows the number of nodes and also the number of GPUs.

Figure 1b shows the speed-up for the multi-GPU and multi-node/multi-core algorithms. It should be clarified that for the calculation of the speed-up (Time(Sequential)/Time(Parallel)), the execution time of the sequential version on CPU was taken as a reference. From the results obtained it is observed that the Speed-up increases as the number of nodes increases for both cases, for 4 processing nodes the multi-GPU algorithm has a Speed-up of 159.428x and 87.057x for the multi-node/multi-core algorithm. The detailed values are expressed in Table 1 along with processing time values in seconds of the algorithms.

5 Conclusions

This paper proposes an implementation on two types of multi-GPU and multi-node/multi-core platforms, for massive searches on unstructured databases. The presented method aims to reduce the time involved in the search process when faced with a large number of simultaneous queries on the system.

Table 1: Processing time (Tp) in seconds and speed-up (Sp) of multi-GPU and multi-node/multi-core algorithms.

Nodes/GPU	muti-GPU		multi-node/multi-core	
	Tp	Sp	Tp	Sp
1	77,097	43,893	134,974	25,071
2	39,344	86,011	68,045	49,732
3	39,687	85,267	49,530	68,322
4	21,226	159,428	38,871	87,057

Experimental validation shows that the multi-GPU approach is 1.6 times superior to the multi-node/multi-core algorithm in solving 56 simultaneous queries. Moreover, in both algorithms the speedup is directly proportional to the number of nodes (number of GPUs) reaching 156x for 4 GPUs and 87x in the case of the hybrid multi-node/multi-core algorithm.

In future work, we plan to increase the number of elements in the database, with the goal of reaching 16 million individuals. Using a database of this size brings with it the problem of the overall memory capacity of the GPU, being necessary to explore a multi-node/multi-GPU approach. Also, we will explore others distribution strategies for the multi-core platform algorithm.

References

1. Barrientos, R.J., Gómez, J.I., Tenllado, C., Prieto Matias, M., Marin, M.: Range query processing on single and multi GPU environments. Computers & Electrical Engineering 39(8), 2656–2668 (Nov 2013), http://www.sciencedirect.com/science/article/pii/S0045790613001560
2. Barrientos, R., Gómez, J., Tenllado, C., Prieto, M., Marin, M.: knn query processing in metric spaces using gpus. In: 17th International European Conference on Parallel and Distributed Computing (Euro-Par 2011). Lecture Notes in Computer Science, vol. 6852, pp. 380–392. Springer, Bordeaux, France (September 2011)
3. De Battista, A., Pascal, A., Gutiérrez Retamal, G.A.: Un nuevo índice métrico-temporal: el historical FHQT (2007), http://hdl.handle.net/10915/22138
4. Gil-Costa, V., Marin, M., Reyes, N.: Parallel query processing on distributed clustering indexes. Journal of Discrete Algorithms 7(1), 3–17 (Mar 2009), http://www.sciencedirect.com/science/article/pii/S1570866708000749
5. Guidet, S., Barrientos, R.J., Hernández-García, R., Frati, F.E.: Exhaustive similarity search on a many-core architecture for finger-vein massive identification. Journal of Physics: Conference Series 1702(1), 012012 (Nov 2020), https://doi.org/10.1088/1742-6596/1702/1/012012, publisher: IOP Publishing
6. Hernández-García, R., Guidet, S., Barrientos, R.J., Frati, F.E.: Massive finger-vein identification based on local line binary pattern under parallel and distributed systems. In: 2019 38th International Conference of the Chilean Computer Science Society (SCCC). pp. 1–7. IEEE (2019)
7. Nations, U.: Macrodatos para el desarrollo sostenible | Naciones Unidas, https://www.un.org/es/global-issues/big-data-for-sustainable-development, publisher: United Nations
8. Rosdi, B.A., W.Shing, C., Suandi, S.A.: Finger vein recognition using local line binary pattern. Sensors 11, 11357–11371 (2011)

Some Issues to Consider in the Management of Energy Consumption in HPC Systems with Fault Tolerance

Marina Morán[1][0000−0002−6334−1190], Javier Balladini[1][0000−0002−9769−7830], Dolores Rexachs[2][0000−0001−5500−850X], and Enzo Rucci[3][0000−0001−6736−7358]

[1] Universidad Nacional del Comahue, Neuquén, Argentina
{marina,javier.balladini}@fi.uncoma.edu.ar
[2] Universitat Autònoma de Barcelona, Bellaterra, España
dolores.rexachs@uab.es
[3] Universidad Nacional de la Plata - CIC, La Plata, Argentina
erucci@lidi.info.unlp.edu.ar

Abstract. Inquiring about different ways to reduce energy consumption during the execution of large-scale applications is essential to maintain and increase the enormous computing power achieved in HPC systems. Fault tolerance methods can have an impact on power consumption. In particular, rollback-recovery methods using uncoordinated checkpoints prevent all processes from re-executing in the event of a failure. In this context, it is possible to take actions on the nodes of the processes that do not re-execute to reduce energy consumption. In this work, we describe some issues to consider when we extend the application of energy-saving strategies beyond the nodes that communicate directly with the failed one.

Keywords: Energy consumption · Fault tolerance · Uncoordinated checkpoints · HPC.

1 Introduction

The energy consumption of supercomputers and HPC systems continues to be a central topic of research. Inquiring about different ways to reduce energy consumption during the execution of large-scale applications is central to maintaining and increasing the enormous computing power achieved.

In parallel HPC message-passing applications, it is mandatory to use some fault tolerance method, which ensures the progress and completion of the application. The most widely used method today is rollback recovery, using coordinated checkpoints. With this method, the application stops every certain period, all the processes perform the checkpoint in a coordinated way and then continue executing. When there is a node failure, all nodes must roll back and resume execution from the last checkpoint. Although this method is widely used due to its low complexity, it is also true that it can imply great pressure on the storage

system due to the simultaneous access of all the processes. A method that relaxes this limitation is the method known as uncoordinated or semi-coordinated checkpoints. In this method, the processes checkpoint independently, and consistency when recovering from a failure is achieved through some support, such as the message log. This method can reduce concurrent accesses to the file system. In addition, only the processes of the node that has failed must be restarted, which prevents all resources perform duplicate tasks. In previous work [6], we evaluate a series of strategies that can be applied to improve energy efficiency when a failure occurs, considering uncoordinated checkpoints. The strategies use the Advanced Configuration and Power Interface (ACPI), in particular we consider the processor P-State that uses the dynamic voltage and frequency scaling (DVFS) techniques and system sleeping states (system states S1 to S4), that is, system hibernation at the node level. By having a characterization of the energy consumption required to execute the application, as in [5], and its communication pattern, we estimate the execution and waiting times of the processes that do not fail. Then, by using a simulator that we have designed and developed, we can evaluate the use of the strategies.

By analyzing the wave of failure propagation, the number of nodes that receive the application of the strategies can be increased. We call this *cascade analysis* and it seeks to extend the application of strategies beyond the processes/nodes that communicate directly with the node where the failure occurred. The greater the number of nodes that receives some strategy, the greater the energy savings. In this work, we consider some aspects to take into account when analyzing cascade blocks.

The rest of the article is organized as follows: Section 2 presents some related works; Section 3 describes the strategies proposed to achieve energy savings, and Section 4 discusses some aspects to take into account in the cascade analysis. Finally, Section 5 presents the conclusions and future work.

2 Related Work

Some works have studied how to take advantage of the waits of processes that do not roll back when a failure occurs. In [1], the authors seek to improve the efficiency of the computer system by replacing the application when the waits are long enough. In [3], the authors simulate a load of an HPC system to evaluate the energy savings when activating and deactivating nodes according to the computational and power requirements of the cluster. Other works slow down the non-critical path to consume less power without substantially increasing execution time [4,7]. [2] can be considered the most similar proposal to this work, since they propose a localized rollback based on the data flow, and reduce the clock frequency of the waiting processes to the minimum possible. We evaluate other strategies, in addition to changing to the minimum frequency, and we do so both for the computation and waits of the processes that continue to execute.

3 Energy-saving Strategies

A series of strategies are defined to be applied to the surviving nodes after the failure, in a parallel message-passing application, running in a homogeneous cluster, which uses the uncoordinated checkpoint fault tolerance method. The failures considered are permanent node failures which in environments using MPI are fail-stop. Fig. 1 shows two processes, P1 and P2, running on different nodes. The green area means regular execution of the application; the red area means that the process is blocked for communication; the blue area means re-execution caused by the failure; while yellow squares indicate that the process is performing checkpoints. When the node where P1 is running fails (indicated by the yellow star), P2 will be affected by the failure because it will have to wait for P1 for a proportional time to its recovery one (recovery time is indicated in blue in the figure). At this time, the strategies to be applied to the computing and waiting phase (indicated in the figure) of surviving process P2 are evaluated.

The strategies are: (1) frequency change for the computational phase, (2) frequency change for the waiting phase, and (3) sleeping for the waiting phase. When the selected strategy is to change the clock frequency, it is applied to all cores of the node.

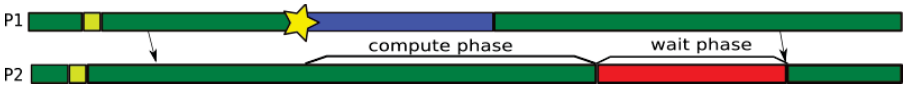


Fig. 1. Computing and waiting phases

4 Cascading Analysis Considerations

In the past, we have developed an event-based simulator to assess the different scenarios that involve the application of the strategies [6]. For the selection of the strategy to apply to each node, it is necessary to know the time until processes of the node get blocked (computing phase) and the duration of that blocking (waiting phase). The simulator results show us that, in general, the duration of the waiting phase defines the action to be applied to each node. If the waiting phase lasts long enough to put the node to sleep, this will be the preferred option, since the savings obtained largely exceed the benefits of the rest. With "long enough to sleep the node" we mean that the wait has a duration greater than the sum of the time it takes for the node to sleep and to wake up plus a configurable fraction of time.

Sleeping a node affects all processes on the node. This is why it is necessary to identify which process will be the one that defines the strategy. In applications with a single process (and multiple threads), there is a single option. In tightly coupled applications, all processes on the node will be blocked nearly at the same

time, and the selected strategy can be applied without a problem. In applications where there are substantial differences in the computing phases of processes that shares the same node, sleeping the node could affect the execution of the other processes.

To which processes is it appropriate to apply the strategy? A first approach is to apply the strategies to the processes that are blocked by communication with the failed process. But it is possible to increase energy savings if we apply the strategies to a larger set of processes that are also affected by the failure, even if they do not communicate directly with the failed process. We then consider analyzing the processes that gets blocked with processes that are already blocked due to the failure. We call this *cascade blocking*. Including this aspect obviously increases the complexity of the algorithm that estimates the computing and waiting phases. Besides, we call *parent process* to the one responsible for blocking the child process. When we analyze processes that are blocked in cascade, some situations may arise:

- The child process communicates several times with its parent before reaching the communication that will block him due to the failure. We define *depth* as the number of subsequent communications to analyze when looking for one that gets blocked due to the failure. The depth can be calculated by looking at the communication pattern between each pair of processes and choosing the maximum amount of communications found (there will never be more communications than that). It would also be possible to define the depth by process, and apply the corresponding one in each case. Here again, we need to check the duration of the recovery. If the recovery time of the failed processes will be greater than a certain threshold, we can ensure that the nodes will sleep. In this case it would not be necessary to consider the depth or make an estimate of the moment when the blocking will occur, since we can mark the action that the node should take, and when that block arrives apply it.
- The parent process has been slowed down (it received the application of a strategy that was to decrease the clock frequency in the computation phase) but the child has not, or it has been slowed down in a different way. This will affect the communication time between the processes, probably generating unexpected waiting pahses.
- If the application uses non-blocking MPI operations, it could happen that a process starts a non-blocking send and goes to sleep (by applying the strategy) before data is transmitted. In this case, when the receiving process issues the matching receive, the communication cannot be completed. But as the node is hibernating, the second process will not be able to complete the communication and will block before expected. A possible solution to this is to wake up the node. In this way, the buffers with the messages become available.

Fig. 2 shows a scenario of 8 processes that communicate in a pipeline manner (each process sends data to the next one) in order to observe the effect of

incorporating the cascade analysis. Fig. (a) shows the case where cascade are not supported, and Fig. (b) shows the case where it is used (depth = 2). The red blocks are waiting times when the node is awake, and the gray blocks are waiting times when the node sleeping. In case (a), the strategies are applied to a single node, while in case (b) the same strategies are applied to the 7 "surviving" nodes. In this way, energy savings increase from 32,500 J to almost 207,000 J. This represents more than six times the number of joules saved in a similar period of time (almost 4 minutes). According to the energy model defined in [6], energy saving is defined as the difference between the joules consumed with and without the application of the strategies. Table 1 shows the results, including cascade with depth = 1. Note that with this depth, only one node is added to the analysis.

Table 1. Selected actions and energy savings

	Compute phase		Wait phase		Sum of the phases			
Node	Action	T (m)	Action	T (m)	T (m)	Save (J)	Save Rate (J/s)	Save (%)
No cascade								
1	No action	4.84	sleep	3.67	8.51	32,502.3	147.77	38.35
Cascade deep 1								
1	No action	4.84	sleep	3.67	8.51	32,502.3	147.77	38.35
2	No action	5.00	sleep	3.68	8.68	32,617.8	147.79	37.74
Cascade deep 2								
1	No action	4.84	sleep	3.67	8.51	32,502.3	147.77	38.35
2	No action	5.00	sleep	3.68	8.68	32,617.8	147.79	37.74
3	No action	5.17	sleep	3.69	8.86	32,764.1	147.82	37.12
4	No action	5.35	sleep	3.70	9.05	32,787.2	147.82	36.39
5	No action	5.52	sleep	3.70	9.22	32,841.1	147.83	35.77
6	No action	5.68	sleep	3.71	9.39	32,925.8	147.85	35.19
7	2.1 GHz	9.22	1.2 GHz	0.35	9.57	10,348.97	18.02	11.08

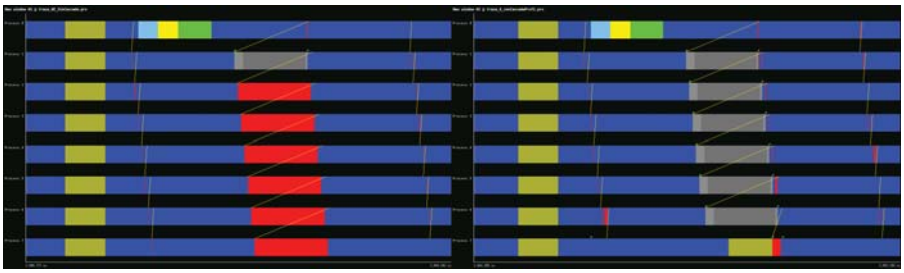


Fig. 2. Application of strategies: (a) Without cascade (b) With cascade

5 Conclusions and Future Work

Energy-saving opportunities exist in a rollback recovery scheme where only some processes must go back and re-execute. If the re-execution time is long and the possibility of sleeping the nodes gets enabled, the energy savings can be large. Depending on the communications pattern, the processes will be blocked sooner or later due to the failure. Extending the application of strategies beyond the nodes that communicate directly with the failed one can improve energy savings, as shown in this work. At the same time, determining which processes block and when they do so, are not easy tasks in applications with loosely coupled and non-homogeneous communication patterns.

Future work includes the analysis of new experimentation that considers non-blocking communications and cascade analysis, and also the implementation of a proof of concept in a cluster.

References

1. Bouteiller, A., Cappello, F., Dongarra, J., Guermouche, A., Hérault, T., Robert, Y.: Multi-criteria checkpointing strategies: Response-time versus resource utilization. In: European Conference on Parallel Processing. pp. 420–431. Springer (2013)
2. Dichev, K., Cameron, K., Nikolopoulos, D.S.: Energy-efficient localised rollback via data flow analysis and frequency scaling. In: Proceedings of the 25th European MPI Users' Group Meeting. pp. 1–11 (2018)
3. Dolz, M.F., Fernández, J.C., Iserte, S., Mayo, R., Quintana-Ortí, E.S.: A simulator to assess energy saving strategies and policies in hpc workloads. In: ACM SIGOPS Operating Systems Review. vol. 46, pp. 2–9. ACM New York, NY, USA (2012)
4. Hajiamini, S., Shirazi, B., Crandall, A., Ghasemzadeh, H.: A dynamic programming framework for DVFS-based energy-efficiency in multicore systems. IEEE Transactions on Sustainable Computing **5**(1), 1–12 (2019)
5. Morán, M., Balladini, J., Rexachs, D., Luque, E.: Prediction of energy consumption by checkpoint/restart in HPC. IEEE Access **7**, 71791–71803 (2019)
6. Morán, M., Balladini, J., Rexachs, D., Rucci, E.: Towards management of energy consumption in hpc systems with fault tolerance. In: 2020 IEEE Congreso Bienal de Argentina (ARGENCON). pp. 1–8. IEEE (2020)
7. Rountree, B., Lowenthal, D.K., De Supinski, B.R., Schulz, M., Freeh, V.W., Bletsch, T.: Adagio: making dvs practical for complex hpc applications. In: Proceedings of the 23rd international conference on Supercomputing. pp. 460–469 (2009)

Big and Open Data

Visual Analytics for Linked Open Data in Marine Sciences

Gustavo Nuñez[1,*], Carlos Buckle[1][0000−0003−0722−0949], and Marcos Zárate[2][0000−0001−8851−8602]

[1] Laboratorio de Investigación en Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (LINVI-UNPSJB), Puerto Madryn, Argentina.
guscostaf@gmail.com cbuckle@unpata.edu.ar

[2] Centre for the Study of Marine Systems, Patagonian National Research Centre (CESIMAR-CENPAT-CONICET), Puerto Madryn, Argentina.
zarate@cenpat-conicet.gob.ar

Abstract The purpose of data exploration and data visualization (DV) is to offer ways of perceiving and manipulating information, as well as extracting and inferring knowledge. In this short paper we present advances in visual representations and intuitive interaction techniques based on artificial intelligence. This contributes significantly to the exploration and understanding of information related to marine sciences represented by ontologies and linked data. This preliminary research allows scientists and non-expert users to analyze sets with information related to oceanography, meteorology and environmental parameters to promote scientific knowledge and productive innovation in the South Atlantic ocean using Linked Open Data (LOD).

Keywords: Data Visualization · Linked Open Data · Marine Science.

1 Introduction

The purpose of DV is to offer forms of information perception and manipulation, as well as knowledge extraction and inference [1,2]. DV provides to users an intuitive way to explore content, identify interesting patterns, infer correlations and causalities, and supports meaning-construction activities. One of the promising approaches to address the problems associated with integration and graphing is to store in a structured way and reproduce the data sets in graphs. The Semantic Web (SW) [3] offers solutions to these needs by enabling LOD [4] where data objects are uniquely identified and the relationships between them are explicitly defined. LOD is a powerful and compelling approach to disseminating and consuming scientific data from various disciplines [5,6,7,8]. It involves publishing, sharing, and connecting data on the Web and offers different methods of aggregation and interoperability.

* Corresponding author.

In recent years, this way of publishing data has been adopted in a large number of LOD disciplines [9], this has made the visualization and exploration of information a crucial task for most of the LOD consumers. Data scientists, domain experts, and citizens want to use intuitive and visual, rather than programmatic, ways to interact with these resources. In the domain of marine sciences, the visualization of data from disciplines such as Oceanography, Meteorology and Biodiversity face great challenges, since there is an exponential increase in their volume due to the growth of technology and the multiplicity of platforms of remote sensing and the demand for knowledge to contribute globally to climate change models [10]. In addition, there is a great diversity in the type of records that must be displayed properly, the physical chemical, geological, meteorological and biological values must be integrated and the analysis/information products must be based on all of them so that the user can make a correct interpretation [11].

The remainder of this short paper is structured as follows: Section 2 presents different initiatives based on LOD for marine sciences. Section 3 presents a proof concept platform developed to visualize information of marine sciences in South Atlantic. Finally, in section 4, we present some conclusions related to previous experiences and planning for future works.

2 Background and Related Work

A large number of LD visualization tools have been introduced in recent years, most of them originating from academia. DV tools in linked data provide graphical representations of a data set or parts of it, with the aim of facilitating its analysis and generating insights from complex, interlinked information on a geotemporal space. Techniques may vary depending on the domain, the type of record, the task the user is attempting to perform, as well as the user's skills.

There are several initiatives carried out in the Argentine context to publish marine science data such as LD, among them we can highlight: [12] which presents the publication of metadata from oceanographic campaigns as LD. OceanGraph [13] defines an oceanographic knowledge graph prototype to manage information from expeditions, scientific publications and environmental variables, while in [14] OceanGraph exploitation is proposed with concrete examples of potential uses by specialists.

At the international level there are also initiatives for the publication of marine science data such as LD, among the main ones we can mention GeoLink [15], a project funded by the EarthCube initiative, which has taken advantage of the principles of LOD to create a database, which allows users to consult and reason in some of the most outstanding geoscience repositories in the United States. The GeoLink dataset includes such diverse information as port calls made by oceanographic cruises, metadata from physical samples, funding of research projects and personnel, and authorship of technical reports. The data has been published in accordance with best practices for LOD [16] and are publicly available through a SPARQL endpoint that currently contains more than 45 million RDF triples.

3 LD Visualization in Marine Science

In the context of marine science, visual exploration is a promising approach for exploring and analyzing data and better understanding the dynamics of complex ocean processes, although publishing data as LD has a number of success stories [15,17], visualization continues to be a problem because it is a particular task that differs from classic DV, mainly due to the characteristics of LD. The use of common vocabularies (cross domains) for the description of the records or the use of typified properties to capture relationships between resources within a set or between different sets, differs from traditional forms of visualization which are unable to capture the complex possible relationships. For the tests described below, we use public information on marine species and environmental variables captured in the South Atlantic through a SPARQL endpoint whose URL is `http://linkeddata.cenpat-conicet.gob.ar/snorql/`. The methodology used for the creation and publication is detailed in [17]

3.1 Case study

Our focus is on the Web-based front-end consisting of querying and visualization tools. We have developed a proof of concept for interactive visualization of oceanographic, environmental and marine biodiversity information. The platform allows the representation and visualization of interactive maps with trajectories of oceanographic vessels, as well as the retrieval of graph schemes of the relationship between environmental variables and species. For which a selection of open source instruments compatible with the visualization of specific types of information was carried out, for example geospatial data, species distribution, traceability and records related to the environment. Figure 1 shows 2 visualizations used to interpret information on a specific species, in this case Mirounga Leonina (Southern elephant seal). Map shows the information of trips made by several individuals during their feeding trips in the sea, additionally overlapping layers with environmental and special information. The other visualization shows bibliographic information associated with the species.

The application is built with the Shiny framework[3] for the R programming language. Access to endpoints is done through the SPARQL package[4]. The application layout is produced with the flexdashboard package[5], and the maps use Leaflet.js, Highcharts, and ggplot2, all accessed through their corresponding R packages.

To see details of the implementation and the source code, see the following link.

[3] `https://shiny.rstudio.com/`
[4] `https://cran.r-project.org/web/packages/SPARQL/SPARQL.pdf`
[5] `http://rstudio.github.io/flexdashboard/index.html`

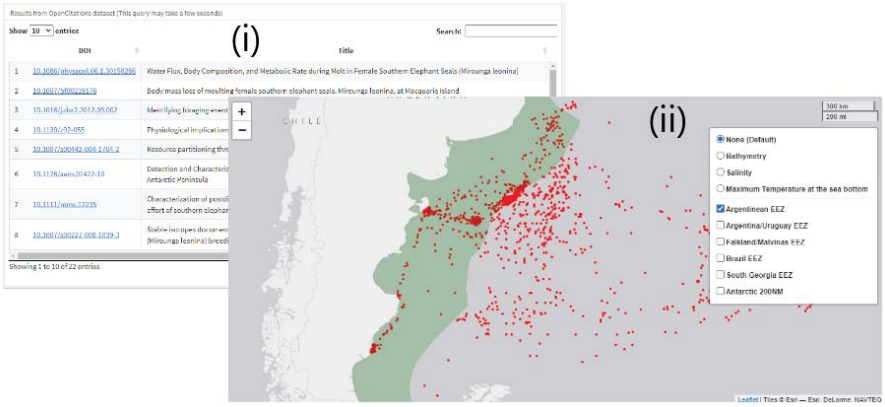


Fig. 1. Visualizations used to relate: (i) marine species with bibliographic information (ii) geo-spatial information of species with environmental variables and marine regions.

4 Conclusions and Future Work

From previous experiences, we can conclude that it is necessary to develop systems capable of visually managing information for comprehensive and secondary use, both from the participating groups and from external users who require information. The results of this preliminary research constitute a substantial contribution, not only for marine sciences, but also as a methodological contribution to scientific visualizations using LD.

As future work, we propose the need to formalize the proof of concept, for this it is necessary to delve into the following aspects: a) Study and research of scientific visualizations typical of marine sciences. b) Develop an online data visualization platform based on prediction models to provide visual analytic facilities and allow interactive queries and analysis of different layers of information. c) Expand the platform or scale the results to other marine spaces, in particular to the Priority Geographic Areas (AGP) of the Pampa Azul initiative[6].

Acknowledgments: This research received funding from project *Linked Open Data Platform for Management and Visualization of Primary Data in Marine Science*. Project No. PI-1562. Financed by Secretariat of Science and Technology of the National University of Patagonia San Juan Bosco (UNPSJB).

References

1. Jeffrey Heer and Ben Shneiderman. Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4):45–54, 2012.

[6] https://www.pampazul.gob.ar/areas-prioritarias/

2. Stratos Idreos, Olga Papaemmanouil, and Surajit Chaudhuri. Overview of data exploration techniques. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 277–281, 2015.
3. Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.
4. Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*, pages 205–227. IGI Global, 2011.
5. Richard K Lomotey and Ralph Deters. Terms extraction from unstructured data silos. In *System of Systems Engineering (SoSE), 2013 8th International Conference on*, pages 19–24. IEEE, 2013.
6. Syed Ahmad Chan Bukhari, Mate Levente Nagy, Paolo Ciccarese, Michael Krauthammer, and Christopher JO Baker. icyrus: A semantic framework for biomedical image discovery. In *SWAT4LS*, pages 13–22, 2015.
7. Syed Ahmad Chan Bukhari. *Semantic enrichment and similarity approximation for biomedical sequence images*. PhD thesis, University of New Brunswick (Canada), 2017.
8. Roderic D.M. Page. Ozymandias: a biodiversity knowledge graph. *PeerJ*, 7:e6739, April 2019.
9. The open linked data cloud. https://lod-cloud.net/, 2021. [Online; accessed 3-May-2021].
10. Tanu Malik and Ian Foster. Addressing data access needs of the long-tail distribution of geoscientists. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 5348–5351. IEEE, 2012.
11. Alex Hardisty and Dave Roberts. A decadal view of biodiversity informatics: challenges and priorities. *BMC ecology*, 13(1):16, 2013.
12. Marcos Zárate, Pablo Rosales, Pablo Fillottrani, Claudio Delrieux, and Mirtha Lewis. Oceanographic data management: Towards the publishing of pampa azul oceanographic campaigns as linked data. In *Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW 2018)*, 2018.
13. Marcos Zárate, Pablo Rosales, Germán Braun, Mirtha Lewis, Pablo Rubén Fillottrani, and Claudio Delrieux. Oceangraph: Some initial steps toward a oceanographic knowledge graph. In Boris Villazón-Terrazas and Yusniel Hidalgo-Delgado, editors, *Knowledge Graphs and Semantic Web*, pages 33–40, Cham, 2019. Springer International Publishing.
14. Marcos Zárate, Carlos Buckle, Renato Mazzanti, Mirtha Lewis, Pablo Fillottrani, and Claudio Delrieux. Harmonizing big data with a knowledge graph: Oceangraph kg uses case. In Enzo Rucci, Marcelo Naiouf, Franco Chichizola, and Laura De Giusti, editors, *Cloud Computing, Big Data & Emerging Topics*, pages 81–92, Cham, 2020. Springer International Publishing.
15. Michelle Cheatham, Adila Krisnadhi, Reihaneh Amini, Pascal Hitzler, Krzysztof Janowicz, Adam Shepherd, Tom Narock, Matt Jones, and Peng Ji. The geolink knowledge graph. *Big Earth Data*, 2018.
16. Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, and Charles Vardeman. Five stars of Linked Data vocabulary use. *Semantic Web*, 5(3):173–176, 2014.
17. Marcos Zárate, Germán Braun, Mirtha Lewis, and Pablo Fillottrani. Observational/hydrographic data of the south atlantic ocean published as lod. *Semantic Web*, 13(2):133–145, 2022.

Big Data and Open Data in Education, evaluation of the scope of existing initiatives. Case study Faculty of Technology and Applied Sciences of the National University of Catamarca

Claudia M. Herrera[1], Carolina I. Chayle[1], Ana C. Pauletto[1], María A. Barrera[1], and Nadir Morales Apaza[1]

[1] Department of Computer Science, School of Technology and Applied Sciences, Universidad Nacional de Catamarca, Argentina. https://tecno.unca.edu.ar
herrera.claudia.mabel@gmail.com

Abstract. Data is the fuel of the 21st century. Increasingly our economies are driven by this input. Data has the potential to create social value, something that higher education institutions are beginning to rely on, as evidenced by the initiatives of public access to their data and the various advantages and benefits provided by big data in improving educational management, the development of customized curricula, monitoring the academic performance of students, as well as in the generation of digital repositories that are the product of years of academic, teaching and research activity. The present research aims to evaluate the scope of existing initiatives on the subject of Big Data and Open Data in education, in the Faculty of Technology and Applied Sciences of the National University of Catamarca (UNCA), producing an effective contribution with preliminary results for the research project called "ICT in the Service of Open Data: Current Situation, Conceptualization and Initiatives for Opening Public Information" which is being carried out in the Department of Informatics.

Keywords: Big Data · Open Data · Education · UNCA

1 Introduction

In recent years, an information revolution has begun worldwide that is directly related to free access to data. Open Data responds to an initiative already installed in several countries and that pursues the publication of data sets that governmental organizations, companies, institutions, among others, have in their possession, so that they can be disclosed and potentially reused by society in general, and in particular, by entities that can add value to such data and thus develop beneficial products for the community.[1]

However, it is impossible to process and analyze huge data repositories with conventional database and analytical tools. This is only possible with the use of the Internet and tools that allow the collection of information, this massive data analysis

is known as Big Data.

The concepts of big data and open data are related, but they are not the same. "Big data" or big data refers to such a massive set of data that it generally requires specialized software to analyze it and extract patterns. In contrast, "open data" refers to the possibility that any person or company can reuse a data set to analyze trends, generate new applications or create businesses, among other uses. For this to happen, data must be accessible in two complementary ways: it must be available and in a format that allows it to be digitally reused. In other words, while big data is defined by its quantity, open data is defined by its quality. Open data does not necessarily have to be "big data" to have an impact. [2]

In addition, the application of Big Data in the field of education requires much deeper research that can provide a broad scope of understanding about the functioning and abilities of higher education institutions in their four primary missions: research and data analysis, teaching, training and administration; through these elements, for example teachers can customize each of their lectures to improve the performance of their students. [3]

This research aims to evaluate the scope of existing initiatives on the subject of Big Data (BD) and Open Data (OD) in education, within the Faculty of Technology and Applied Sciences (FTyCA) of the National University of Catamarca (UNCA), producing an effective contribution with preliminary results for the research project called "Las Tic al Servicio del Dato Abierto: Situación Actualización, Conceptualización e Iniciativas de Apertura de Información Pública" which is executed in the Department of Informatics.

2 Big Data

Although there is no commonly accepted definition of DBs, we can say that they are data that can be defined by a combination of the following five characteristics [4]:

- *Volume:* where the amount of data to be stored and analyzed is large enough to require special considerations.
- *Variety:* where the data consists of multiple types of data, potentially from multiple sources; here we should consider structured data stored in tables or objects for which metadata is well defined, semi-structured data such as documents or other similar where metadata is contained internally (e.g. XML documents) or unstructured data, which may be photographs, video or any other form of binary data.
- *Velocity:* where data is produced at high rates and running on "stale" data is not valuable.
- *Value:* when the data has a perceived or quantifiable benefit to the company or organization using it.
- *Veracity:* where the accuracy of the data can be assessed.

All these features today are undoubtedly provided by new technological trends, such as cloud computing and the Internet, without the latter this would not be possible. The versatility and magnitude with which data is handled through these tools facilitates and shortens the distances of time and space, which in itself is a great benefit for educational institutions and the social impact before a society in constant

advance. The benefits of using DB techniques are quite broad. Two main groups of benefits emerge: cost savings; and competitive advantage. The potential of BD depends on the sectors of activity where it is used. There have been significant advances in science through the adoption of BD, particularly in astronomy, biology and bioinformatics. [4]

3 Open Data

OD (Open Data) are those data that are freely available for use, reuse and redistribution. These data must comply with the following premises [5]:

- *Availability and access:* Data should be available at a reasonable cost and in a convenient and modifiable form.
- *Reuse and redistribution:* The format of the data will allow for reuse, redistribution and integration to other datasets.
- *Universal participation:* Everyone should be able to use, reuse and redistribute information without restrictions.

In December 2007, the Opengov Data organization published eight principles that should be considered by governments when publishing data. These principles do not indicate which data should be public, but rather specify the conditions for public data to be open. [5]:

- *The data must be complete. All data shall be available.* Data is stored electronically including documents, databases, transcripts, audio and video recordings, images, etc. Public data is data that is not subject to existing privacy, security or privilege limitations, which is governed by other laws.
- *Data should be primary.* Data are published as extracted from the source, at the best possible level of granularity, not in aggregated or modified forms.
- *Data must be timely.* Data are made available as soon as necessary to preserve the value of the data.
- *Data must be accessible.* Data are available to the widest range of users and for the widest range of purposes.
- *The data shall be machine processable.* The data shall be reasonably structured to permit automatic processing of the data.
- *Access should be non-discriminatory.* The data are available to anyone, without registration.
- *The format of the data should not be proprietary.* The data are available in a format over which no entity has exclusive control.
- *The data must be license-free.* The data are not subject to any copyright, patent, trademark or trade secret regulations. Reasonable privacy, security and privilege restrictions may be allowed as they are governed by other laws.

Tim Berners-Lee proposed a categorization of the degree of openness of data based on stars, regarding how open and usable data an institution can offer (see Fig.1). [6]

- *1 star:* Data must be available on the web in any format and under an open license to be considered open data.
- *2 stars:* Data must be structured and in a format that can be interpreted or

processed by machines.

- *3 stars:* Same as above but in a non-proprietary format.
- *4 stars:* All of the above plus the use of w3c recommended standards (RDF - SPARQL) to identify things.
- *5 stars:* All of the above plus linking to the data of others and thus providing context.

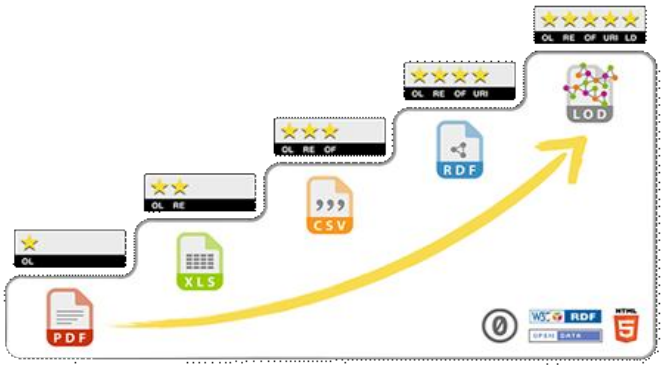


Fig. 1. 5-star development scheme for open data.

4 Big Data and Open Data in Education

BD in the educational world refers, above all, to the enormous amount of data and information we have about the educational system itself, the academic performance of students, the training needs of teachers, or the impact of educational measures taken in the field of training to improve the quality of educational processes in times of globalization, interculturalism and multilingualism. Thus, one of the key features of BD in the academic context is linked to learning analytics, which are understood as processes of continuous and procedural assessment that address the changes and evolution of certain school factors or variables. Thus, the potential of BD can be specified in the following aspects [7] Among the main educational methods derived from BD analytics (whose origins date back to the year 2000), and its integration with the new smart devices and web technology that today are already being applied in the educational field, are the following [8]:

- *Adaptive learning:* knowledge data, which will facilitate analytics to create the tailored curriculum. This method emphasizes the areas in which students have more difficulties, in order to adapt to their way and pace of learning and create a personalized, differentiated and tailored teaching path for each student.
- *Competency-based education:* today's competency-based education platforms, derived from the power of DB analytics under learning metrics, empower student independence while allowing teachers to assess progress. One example is Mastering.
- *Flipped classroom and blended learning:* A typical platform that uses these educational methods is Moodle. The knowledge bases generated by these types of

platforms serve to complement the data lakes of an educational institution and thus facilitate learning analytics.

- *Gamification:* This tool helps teachers enhance classroom learning in an efficient and fun way. In particular, it enriches specific behaviors and boosts students' active participation by issuing rewards in real time. This platform delivers tracking reports to inform students' progress to both parents and teachers.
- *Mobile learning:* Currently, these devices promote learning through educational content, such as electronic textbooks or interactive courses that help improve school performance. They also allow the application of assessment or placement tests that speed up and simplify the student evaluation process. These devices are one of the most important pillars for gathering information about students and the educational processes they attend, thus increasing the possibilities of performing analytics with BD technology and identifying new areas of opportunity to improve education.

These methods, which allow education to reach a larger student population and meet the continuous demands of students (who are increasingly demanding a more flexible and personalized education), are providing very promising results and have changed the way in which students relate to each other, to their teachers and to educational institutions.

For their part OD can be used in higher education, as follows [9] :

- Collaborating with researchers on real research projects.
- Promoting collaboration among students from diverse disciplines by creating learning activities based on sustainable development goals.
- Encouraging activities in which students help their local communities solve real-life problems.

Likewise, the use of OD can contribute to the development of critical thinking, research, teamwork and citizenship skills. So, when working with students, we can incorporate OD into teaching activities by guiding learning through [9]:

- Identify and describe the learning outcomes for the planned activities.
- Identify the portals that will generate the data.
- Identify and clearly describe the challenges that students might face.
- Provide training materials for the software that students will need to analyze the data.
- Support students in communicating their findings to local communities.

5 Methodology

The research is within the paradigm of non-experimental research design, of the exploratory type, observing the phenomenon as it occurs without manipulation of variables or comparison of groups of subjects. [10]. A descriptive and comparative study was carried out. A search strategy was proposed considering the characteristic aspects of each of the terms related to the object of study, i.e., BD and OD in education that allow the evaluation of existing initiatives in the Faculty of Technology and Applied Sciences (FTyCA) of the National University of Catamarca (UNCA). With respect to obtaining information, samples were taken through direct observation

of the FTyCA website.

6 Preliminary results

The FTyCA showed a considerable amount of data published in the different domains that articulate the university (institutional, academic, teaching, research repositories, archives). Specifically, the following were observed: a research repository, publication of government and academic acts, among others. In other words, actions corresponding to the first steps in OD were evidenced when opening small and simple data sets, which are part of a larger data set. All published data have an open format but the level of openness corresponds to 1 star. This is because most of the published data are in a format that does not allow their reuse, for example in .PDF, .DOC, .ODT, .TXT, among others; or a scanned image. Furthermore, these data are available free of charge without the need to register or request access. On the other hand, only some of the data published can be downloaded from the URL where they were found and comply with the "Updated Data" evaluation index, which guarantees users that the data are renewed every year.

In relation to the observation and exploration of the aforementioned aspects of BD, the FTyCA has the Moodle virtual education platform in which applications of competency-based education, flipped classroom and blended learning, and gamification were found. It also offers teachers and students the possibility of applying Mobile Learning through the application of this platform, which is also offered by the academic unit. Although different statistics are available on student learning in terms of the number of passing and failing students, student desertion, enrollment dropout, among others, no evidence was found of the application of BD for adaptive learning in customized study plans that take into account these statistics, learning habits, knowledge, weaknesses and strengths of each student.

7 Conclusions

Based on the study conducted and the results obtained, the following comments can be made:

- To advance in the opening of data it is necessary to consider the criteria for selecting data for opening provided by the Undersecretary of Public Innovation and Open Government of the Ministry of Modernization of the Nation in the "Open Data Kit".
- Note that data openness focuses on non-personal data that are not subject to specific restrictions.
- To give greater impetus to the use of big data in the field of FTyCA.
- To train human capital for the application of big data technology in the academic unit.

Open data is important for the university because it encourages greater transparency in the information it produces and efficiency in its actions. For society,

these open data are important because they provide information about the different activities carried out in the university environment as well as the careers taught, the number of students, the number of graduates, research, etc.. Likewise, they allow to know the current problems such as student desertion and which are the actions that are being carried out to solve them.

It would also be very beneficial for the Faculty of Technology and Applied Sciences to take advantage of the benefits offered by Big Data to move from the massive collection of data and the mere generation of statistics, to the positive and integrative use of educational information to achieve greater quality and management of the education system, and achieve academic excellence; To this end, the curricula can be adapted with the knowledge of the strengths and weaknesses of the students, through the statistics produced, in order to reduce dropout and dropout rates; to bring students closer to employment; and to produce research and contributions to the local public and private sector according to the needs.

References

1. Gobierno de Chile, Unidad de Modernización y Gobierno Digital, Minsegpres. Norma Técnica para Publicación de Datos Abiertos en Chile. 2013.
2. Straface Fernando. Big data y datos abiertos: cómo crean valor gobiernos y empresas. Cronista, 2019. https://www.cronista.com/columnistas/Big-data-y-datos-abiertos-como-crean-valor-gobiernos-y-empresas-20190310-0026.html#:~:text=Es%20decir%2C%20mientras%20que%20big,data%22%20para%20generar%20un%20impacto. Visitado última vez 14 de abril de 2022.
3. Heredia Mayorga, H. Big Data en la educación. Artículo de Investigación. Horizontes. Revista de Investigación en Ciencias de la Educación. Volumen 4 / No. 16. ISSN: 2616 - 7964. ISSN-L: 2616 - 7964. pp. 523 - 535. Octubre - Diciembre 2020. https://doi.org/10.33996/revistahorizontes.v4i16.134
4. Tejada-Escobar, F., Murrieta-Marcillo, R., Villao-Santos, F., Garzón-Balcázar, J. Big Data en la Educación: Beneficios e Impacto de la Analítica de Datos. Revista Científica y Tecnológica. UPSE. Vol. 5, Nº 2 (2018), 88-96 (Enero – Junio 2019).
5. AGESIC - Agencia para el Desarrollo del Gobierno de Gestión Electrónica y la Sociedad de la Información y el Conocimiento. Datos Abiertos de Gobierno. Guía Básica de Apertura y de Reutilización. Buenas Prácticas. Versión 1.0 – 2012.
6. Tim Berners-Lee. 5 Estrellas Datos Abiertos. https://5stardata.info/es/ visitado última vez el 14 de abril de 2022.
7. Matas, A., Leiva, J.J. & Nuñez, N. El Big Data desde la perspectiva de una muestra de estudiantes de universidades del norte de Perú. Propósitos y Representaciones, 8(1), e376. 2020.
8. Salazar Argonza, J. Big data en la Educación. Revista Digital Universitaria. UNAM. Vol. 17, Núm. 1. ISSN 1607 - 6079.2019.
9. Atenas, J., Raffaghelli, J. Los Datos Abiertos en la didáctica universitaria. Herramientas para el desarrollo del pensamiento crítico. Workshop Red Argentina de Educación Abierta. 2020.
10. Rivas-Rebaque B., Gértrudix-Barrio F., Cisneros de Britto J.C. La percepción del docente universitario ante el uso y valor de los datos abiertos. Educación XX1, vol. 22, núm. 2, pp. 141-163. Universidad Nacional de Educación a Distancia. España. 2019.

Artificial and Computational Intelligence

A Preliminary Approach about using nowadays Knowledge Engineering in Artificial Intelligence: a literature overview

Cinthia Vegega[1[0000-0002-5382-7875]], María F Pollo-Cattaneo[1[0000-0003-4197-3880]] and José A Olivas[2[0000-0003-4172-4729]]

[1] Grupo de Estudio de Metodologías para Ingeniería en Software (GEMIS)
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Argentina
[2] Soft Management of Internet and Learning (SMILe)
Universidad de Castilla-La Mancha, España
cinthia.vegega@gmail.com, flo.pollo@gmail.com,
joseangel.olivas@uclm.es

Abstract. This paper presents a first literature overview on the nowadays use of Knowledge Engineering in the framework of the development of solutions in the field of Artificial Intelligence. With the assumption that the conceptualization, formalization and modeling of data is fundamental in this type of projects, it is considered that Knowledge Engineering can actively collaborate orienting and guiding part of these activities. In this context, a doctoral research line is proposed about the use of Knowledge Engineering to perform Intelligent Data Analysis. This line of research arises from the fact that existing methodologies in the field of Artificial Intelligence and Data Mining do not incorporate complete domain and context knowledge in data analysis. The incorporation of this knowledge allows contrasting the hypotheses obtained from the data and enriches the analysis reaching better results.

Keywords: Knowledge Engineering, Intelligent Data, Artificial Intelligence

1 Introduction

Knowledge Engineering [1,2] is a discipline that is linked to Artificial Intelligence and is oriented to the construction of Intelligent Systems, which are artifacts that present some intelligent behavior in the human sense. Within Intelligent Systems are Knowledge-Based Systems whose source of knowledge can come from data or domain knowledge. The data may be obtained from the domain itself or could be external, so they may require data from libraries or other databases outside the domain.

Data analysis methods typically use normalized and structured numerical data as inputs, without considering the heterogeneous nature of the data repositories. These methods try to interpret the output, but a complex Intelligent System cannot be analyzed through them. If added to this data analysis, domain knowledge would be enriched and

complement the development of Intelligent Systems. In this sense, Knowledge Engineering can help with the work of data that may be inaccurate, uncertain, have errors, contradictions and biases. If the data is biased, there is a risk of generating an Intelligent System that is not based on reality and produces erroneous results [3], solving a different problem from the one that is to be solved.

Also, the usual practice to find relationships between existing data involves "going blind" in their processing using the application of algorithms or analysis tools, such as statistical tools or Machine Learning [4]. This is done in a unidirectional way with the purpose of being able to interpret the output generated by these algorithms in order to verify if the dataset can be useful or not for the implementation of an Intelligent System. However, this practice is not valid for an Intelligent System due to its complex nature.

In this context, the doctoral research line proposes that, based on existing data and domain knowledge, the process of Intelligent Data Analysis can be systematized using techniques and tools belonging to Knowledge Engineering. This analysis will focus on different elements such as the contrast and establishment of hypotheses between data and domain knowledge, the bidirectional balance between correlation and causation, searching for association rules in both directions, and the construction of formalizable cognitive models, such as taxonomies and ontologies, that enrich the analysis.

2 Motivation and literature overview

Hybrid Intelligence [5,6] allows achieving complex objectives by combining human intelligence and Artificial Intelligence in order to achieve results superior to those that each of them could have achieved separately and to continuously improve by learning from each other. In this sense, the combination of Knowledge Engineering and Machine Learning allows complementing the strengths and weaknesses of both, opening new possibilities for the organization of knowledge and contributing to the creation of hybrid Intelligent Systems. This is why, currently, there are international forums that are beginning to address the importance of the relationship between Knowledge Engineering and Machine Learning, such as AAAI-MAKE [7], KEPS [8], IKE [9] and KEOD [10].

Some works that show the application of Hybrid Intelligence are:

- In [11] a recommendation system for small businesses with Hybrid Intelligence is proposed to prevent the lack of operating funds in the short term that represents a critical problem in this type of organizations. Since the solution using Machine Learning alone was not viable due to the complexity of the problem, the heuristic knowledge of financial experts is integrated. This knowledge is incorporated through a rule-based analysis in order to identify different corrective actions to solve the problem while Machine Learning models are used to predict the probability of success and the cost of each action.

- In [12] a model is presented that combines the results of different agricultural simulations with the knowledge of the experts to make predictions. Artificial Neural Networks are used to make these predictions, while ontologies are used to link the data with the experts' knowledge.
- In [13] a decision support system is designed for the admission process of students to a University in order to admit the right students. Since the current admission process is based on experts who apply their criteria for admitting or not admitting a student, their knowledge was acquired through previous case studies. It combines reasoning based on previous cases and Machine Learning for the analysis of qualifications.

On the other hand, knowledge is usually used anarchically, since there are methodologies such as CRISP-DM [14], SEMMA [15] or P3TQ [16] that define an understanding of the domain, but do not systematize how to use it and do not relate it directly to data analysis. Likewise, the KDD process [17] proposes a data preprocessing and transformation stage to then detect behavioral patterns, but does not incorporate domain knowledge in the data analysis either.

In this line of doctoral research we intend to deconstruct what has been done in the implementation of Intelligent Systems that are currently successful cases, such as the evaluation of hate speech in social networks [18] or INFEDEC 2.1 [19] which is a Knowledge Based System for decision support in the medical field.

In order to make the appropriate decision, the right knowledge needs to be available and accessible in order to be able to make better decisions [17]. When we are faced with Artificial Intelligence to be able to anticipate the future by looking for the right outcome, today, there is a decision of whether we link it to the data that is available or we also incorporate domain knowledge into that data. In this sense, two tasks can be performed from the available data. In the first instance, their manipulation and in the second instance their exploitation.

Data manipulation consists of surveying the available repositories in the domain and from them, being able to transform them into a heterogeneous datalake containing usable data. All available data must be acquired with the objective of converting them into a representation appropriate to the problem being solved. They must be prepared in a certain format and then cleaned so that they can be used and stored in a suitable container.

The objective of *data exploitation* is the complete analysis of the data. Therefore, from the datalake it must be decided whether the data or a part of it follows a trend or not. In the case that part of the data follows a trend, an analysis of the data is performed using statistical models and Machine Learning. However, whether they follow a trend or not, the use of Knowledge Engineering is also proposed in order to contrast and establish hypotheses based on the experience of experts in that area of knowledge. In reference to the contrast and establishment of hypotheses between data and domain

knowledge, Popper [18] defines that in the search for knowledge one cannot dispense with the truth, therefore, it is a matter of finding true theories or those that are close to it, rejecting false theories. The hypotheses that can be established from the data have a degree of confirmation and also have a degree of refutation that can be obtained from the knowledge of the domain, giving more or less credibility to each of them.

Knowledge Engineering allows the acquisition of expert knowledge through direct methods, such as interviews, questionnaires and surveys, or through indirect methods, such as observation of routine tasks, analysis of protocols and grids [19]. This knowledge can be conceptualized taxonomically and ontologically, enriching the data analysis. Soft Computing [20] techniques that are tolerant to imprecision and uncertainty within the knowledge representation can also be used.

2 Conclusions

This article describes a doctoral research line that proposes an Intelligent Data Analysis using techniques and tools belonging to Knowledge Engineering. Knowledge Engineering enriches the data analysis process with domain knowledge. In this way, Machine Learning is used to generate hypotheses that check or contradict with domain knowledge, being able to find a balance between correlation and causation.

As future lines of work and considering that in order to carry out any solution development process in the field of Artificial Intelligence, it would be advisable to be oriented by guidelines, models, methodologies and/or frameworks that guarantee the implementation of a high quality product, it is expected to lay the foundations for the proposal of an integral methodology based on good engineering practices with a view to the fields of Knowledge Engineering and Data Analytics. The proposed framework aims to contribute to strengthen the relationship between specifications and their formal representations in search of a structured and organized implementation of the intervening actors.

References

1. García Martínez, R., Britos, P.: Ingeniería de Sistemas Expertos, Editorial Nueva Librería, 2004.
2. Palma, J. T., Paniagua-Arís, E., Martín, F., Marín, R.: Ingeniería del Conocimiento. De la Extracción al Modelado de Conocimiento, Inteligencia Artificial, vol. 11, pp. 46-72, 2000.
3. Domingos, P: A Few Useful Things to know about Machine Learning. Communications of the ACM, 55(10), 78-87, 2012.
4. Olivas Varela, J. A.: Inteligencia Artificial, Inteligencia Computacional y Análisis Inteligente de Datos, OBS Business School, 2021.
5. Dellermann, D., Ebel, P., Söllner, M., Leimeister J. M., Hybrid Intelligence, Business & Information Systems Engineering 61 (5), 637-64, 2019.
6. Kamar, E., Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence, IJCAI, 4070-4073, 2016.

7. AAAI-MAKE, https://www.aaai-make.info, last accessed 08/04/2022.
8. KEPS, https://icaps21.icaps-conference.org/workshops/KEPS, last accessed 08/04/2022.
9. IKE, https://www.american-cse.org/csce2021/conferences-IKE, last accessed 08/04/2022.
10. KEOD, https://keod.scitevents.org, último acceso 08/04/2022.
11. Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R, CRISP-DM 1.0 Step by step BI guide. Edited by SPSS, https://www.the-modeling-agency.com/crisp-dm.pdf, last accessed 08/04/2022.
12. SAS, Introduction to SEMMA, https://documentation.sas.com, last accessed08/04/2022.
13. Pyle, D., Business Modeling and Data Mining, Morgan Kaufmann Publishers, 2003.
14. Fayyad, U., Piatestky-Shapiro, G. y Smyth, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. Comunications of the acm, 39(11), 27-34, 1996.
15. Montoro Montarroso, A.: Análisis de Sentimientos para la prevención de mensajes de odio en las Redes Sociales, Trabajo Fin de Grado, Universidad de Castilla-La Mancha, Escuela Superior de Informática, 2019.
16. Lago R., Olivas Varela J. A., Megido J. A., Grávalos C., Pérez J.: INFEDEC 2.1: Sistema de Ayuda en la Decisión de Diagnóstico y Tratamiento de Enfermedades Infecciosas. Actas de ESTYLF'96: VI Congreso Español sobre Tecnologías y Lógica Fuzzy, Universidad de Oviedo, pp. 271-275, 1996.
17. Bennet, A., Bennet, D.: The Decision-Making Process in a Complex Situation, In: Handbook on Decision Support Systems 1, pp. 3–20, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
18. Popper, K. R., La Lógica de la investigación científica, Sánchez Zabala, V. (trad.), Madrid: Tecnos, 1962.
19. García Martínez, R., Rossi, B. & Britos, P.: Metodologías de Educción de Conocimiento para la Construcción de Sistemas Informáticos Expertos, Revista del Instituto Tecnológico de Buenos Aires 21:72-84, 1998.
20. Zadeh, L. A: Soft Computing and Fuzzy Logic. En Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi a Zadeh, p. 796-804, 1996.

TinyML for Small Microcontrollers

César A. Estrebou[1][0000−0001−5926−8827], Marcos D. Saavedra[2], Federico Adra[2], and Martín Fleming[2]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata
{cesarest}@lidi.info.unlp.edu.ar
[2] Facultad de Informática, Universidad Nacional de La Plata

Abstract. This paper describes the progress made in the context of a research and development project on machine learning techniques and algorithms applied to small microcontrollers. The beginning of the development of EmbedIA, a machine learning framework for microcontrollers, is presented. The experiments carried out comparing the proposed framework with other similar frameworks such as *Tensorflow Lite Micro*, *µTensor* and *EloquentTinyML* show an important advantage with respect to memory and inference time required by small microcontrollers.

Keywords: Machine Learning · Embedded Systems · Microcontrollers · IoT · Convolutional Neural Networks · TinyML

1 Introduction

While machine learning is a term that encompasses many different approaches to solving problems, TinyML is a subset of machine learning that refers specifically to the application of machine learning on resource-constrained devices like microcontrollers.

In the past few years, the world has been experiencing a real proliferation of new smart devices. One of the most interesting aspects of these is that, for the first time, they are capable of running machine learning models on the device itself. One of the main causes of this evolution is the paradigm shift where all the information that was sent to the cloud for processing, began to be processed at the edge to avoid problems [5, 10] related to bandwidth, response delays, high computational and storage costs, higher energy consumption, among others.

However, machine learning is a complicated discipline and making it work on small devices creates more than interesting challenges. This is why TinyML has become a popular area of research and development, where machine learning and in particular deep learning have the potential to provide powerful solutions [9] as long as they can be adapted to devices with limited hardware resources.

In this context, in 2021 we initiated this research and development project that aims to document, study and implement traditional machine learning techniques adapted to small devices. In this paper we present the progress made during the course of the past year.

2 Software and Hardware for TinyML

2.1 TinyML Microcontrollers

From the hardware point of view, the term TinyML is associated with relatively powerful devices, with significant memory and computing capabilities for what is considered a "traditional" microcontroller. On the other hand, TinyML is closely related to the IoT area, so it is usually considered only devices with wireless communication.

Within the project, all microcontrollers with a minimum capability to run machine learning algorithms are included for experimentation. The choice of models is based on aspects such as local availability, low cost, low/medium computational capacity and availability of open source software for application development. Regarding connectivity, both IoT and non- IoT devices are considered, since from a machine learning point of view there are many popular devices with interesting hardware capabilities without this feature.

Currently, the project has several development boards for testing different implementations of machine learning algorithms. The table 1 shows the microcontrollers that have been tested as well as their technical characteristics.

Table 1: Relevant technical features of the MCUs used in the project.

Board	MCU	Cores	Clock (Mhz)	Memory (KiB)			FPU	Connectivity
				Bits	Data	Prog.		
Arduino Mega	ATmega2560	1	16	8	8	256	No	No
Stm32f103c8t6	Arm Cortex-M3	1	72	32	20	64	No	No
Stm32f411ceu6	Arm Cortex-M4	1	100	32	128	512	Si	No
NodeMCU	Tensilica L106	1	80	32	80	512	No	Wi-Fi
ESP32-WROOM	Xtensa LX6	2	160	32	520	448	Si	Wi-Fi+BT
Raspberry Pi Pico	RP2040	2	133	32	264	2048	No	No

2.2 On-Line Platforms and Open Source Frameworks

There are a variety of on-line platforms (*AlwaysAI*, *Edge Impulse*, *Qeexo*, *Cartesiam.AI* and *OctoML*, among others) that significantly simplify the development and deployment of machine learning applications on microcontrollers. Some perform automatic exploration of solutions for data, others allow the configuration of models from data and others optimize models for deployment on microcontrollers.

Regarding open source libraries and frameworks, there are several alternatives but few of them provide support for neural networks and even fewer support convolutional neural networks. Among those analyzed within the project,

the following can be mentioned *EloquentTinyML* [3], *Tensorflow Lite Micro*[4] [1], *µTensor* [5], *EdgeML* [2] and *CMSIS-NN* [7]. In general, they all require experienced users with respect to the generation of machine learning models as well as the microcontroller development platform.

Both online platforms and open source frameworks/libraries are mostly developed and optimized for specific architectures such as *ARM Cortex-M*, for 32-bit architecture or microcontrollers with support for FPU, DSP and/or SIMD instructions. This excludes devices with different architectures or without hardware for specialized mathematical computing. Another limitation that these libraries usually have is that they are developed for C++ 11 and rely on heavy software architectures, object-based with inheritance and polymorphisms that increase the size of the programs and slow down algorithm inference time. This approach may be feasible for microcontrollers with good memory size and hardware resources that accelerate mathematical computing, but it is unsuitable for microcontrollers with low computing power and limited hardware resources.

2.3 Development of a New Open Source Framework

As mentioned in the section 2.2 the online platforms and frameworks have important limitations. In this context, it was decided to start the development of EmbedIA, an open source framework to implement machine learning solutions on microcontrollers with important hardware limitations. In this first stage it was decided to focus on the implementation of neural networks, giving priority to the layers and functions of convolutional neural networks.

EmbedIA is implemented in C, C++ and Arduino compatible code so that it can be compiled on any platform that supports these programming languages. It provides functionalities to perform inference and debugging of the models from the microcontroller. Currently, it supports different neural network layers and activation functions including convolutional, depthwise, binary, pooling, flatten, fully connected, batch normalization, ReLU, sigmoid and softmax. It integrates optimizations for 32-bit, 16-bit and 8-bit fixed-point arithmetic. This reduces program size and RAM usage and speeds up inference time on microcontrollers without floating-point support.

To improve the performance of data memory usage, a swap buffer is implemented to minimize the amount of dynamic memory requirements and avoid fragmentation, which is indispensable for those microcontrollers that do not implement good memory management.

In addition, a conversion tool is provided to transform a model generated with *Tensorflow/Keras* to its equivalent in C code. It allows to generate a C, C++ or Arduino project that includes the functions to perform the inference on the converted model, possibility to use fixed point or floating point and debugging functions for the model.

[3] https://github.com/eloquentarduino/EloquentTinyML
[4] https://www.tensorflow.org/lite
[5] https://github.com/uTensor/uTensor

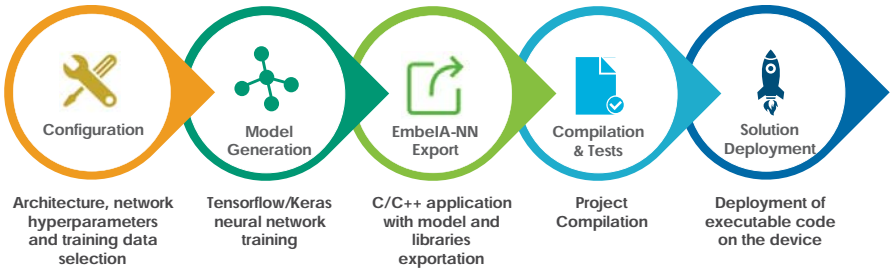


Fig. 1: Steps of development process with the EmbedIA Framework.

3 Experiments Performed and Results Obtained

A series of experiments have been performed on several convolutional neural network [6] (CNN or ConvNet) models to compare the capabilities of EmbedIA with respect to other frameworks/libraries. Initially, we considered to test the performance of *Tensorflow Lite Micro*, *EloquentTinyML*, *CMSIS-NN* and *EdgeML* but, unfortunately, the latter two do not have support for the microcontrollers used.

Among the most relevant results we can highlight [3, 4] one where a convolutional neural network model was used to recognize a total of 36 classes associated with images of 26 letters and 10 handwritten digits. The four Embedia-NN implementations (8-bit, 16-bit and 32-bit fixed-point and floating-point) were compared with other similar frameworks/libraries on microcontrollers of different features, measuring data and program memory size required and inference time consumed. It was found that the four variants of the proposed framework clearly outperformed the implementations of μ*Tensor*, *Google Tensorflow Lite Micro* and *Eloquent TinyML*. In particular the 16-bit fixed-point implementation achieves, on average, a 5 to 10 times improvement in inference time, about 3 times the data memory requirements and 3 to 7 times the program memory requirements.

As part of the exploration of EmbedIA's capabilities we can mention the development of two interesting models that run on different microcontrollers. The first model runs on an 8-bit *Atmega 2560* with 8 Kib of RAM and recognizes 26 handwritten characters. The second runs on a *Tensilica Xtensa LX6* and recognizes 6 different voice commands. It should be noted that both models could not be used in other frameworks because they exceeded the amount of memory of the microcontrollers.

4 Final Comments

This paper has presented the progress of the research and development project on machine learning applied to microcontrollers with significant hardware limitations. It has started the development of EmbedIA, a machine learning frame-

work for microcontrollers, which at this stage is focused on the implementation of neural network algorithms. So far, comparative performance tests with other frameworks show the potential of this new framework.

Currently, the implementation of convolutional layers for binary neural networks is being finalized and experiments are being prepared to compare them with the implementation of traditional convolutional networks. Different convolutional models are also being developed and tested for human detection on microcontrollers with a small camera.

In the future, it is planned to implement a convolutional layered optimization using an optimization called *patch-based interference* [8] which minimizes about 8 times the peak RAM requirement.

References

1. David, R., Duke, J., Jain, A., Reddi, V.J., Jeffries, N., Li, J., Kreeger, N., Nappier, I., Natraj, M., Regev, S., Rhodes, R., Wang, T., Warden, P.: Tensorflow lite micro: Embedded machine learning on tinyml systems (2021)
2. Dennis, Don Kurian and Gopinath, Sridhar and Gupta, Chirag and Kumar, Ashish and Kusupati, Aditya and Patil, Shishir G and Simhadri, Harsha Vardhan: EdgeML: Machine Learning for resource-constrained edge devices, https://github.com/Microsoft/EdgeML
3. Estrebou, C.A., Fleming, M., Saavedra, M.D., Adra, F.: A neural network framework for small microcontrollers. In: Proceedings of the XXVII Argentinean Congress of Computer Science. pp. 51–60 (2021)
4. Estrebou, C.A., Fleming, M., Saavedra, M.D., Adra, F., De Giusti, A.E.: Lightweight convolutional neural networks framework for really small tinyml devices. In: Narváez, F.R., Proaño, J., Morillo, P., Vallejo, D., González Montoya, D., Díaz, G.M. (eds.) Smart Technologies, Systems and Applications. pp. 3–16. Springer International Publishing, Cham (2022)
5. Farhan, L., Kharel, R., Kaiwartya, O., Quiroz-Castellanos, M., Alissa, A., Abdulsalam, M.: A concise review on internet of things (iot) -problems, challenges and opportunities. In: 2018 11th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP). pp. 1–6 (July 2018). https://doi.org/10.1109/CSNDSP.2018.8471762
6. Goodfellow, I.J., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge, MA, USA (2016), http://www.deeplearningbook.org
7. Lai, L., Suda, N., Chandra, V.: Cmsis-nn: Efficient neural network kernels for arm cortex-m cpus (2018)
8. Lin, J., Chen, W.M., Cai, H., Gan, C., Han, S.: Mcunetv2: Memory-efficient patch-based inference for tiny deep learning. In: Annual Conference on Neural Information Processing Systems (NeurIPS) (2021)
9. Sharma, K., Nandal, R.: A literature study on machine learning fusion with iot. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). pp. 1440–1445 (April 2019). https://doi.org/10.1109/ICOEI.2019.8862656
10. Shekhar, S., Gokhale, A.: Dynamic resource management across cloud-edge resources for performance-sensitive applications. In: 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. pp. 707–710 (May 2017). https://doi.org/10.1109/CCGRID.2017.120

An analysis on how can AI empower the senior population in their access to banking services

Yolanda Alfaro Peral[1], Eugenio Concepción[2], Izaskun López-Samaniego[1], and Gonzalo Zarza[1][0000−0003−0795−0933]

[1] Design and Data & AI Studios, Globant, Spain
{yolanda.alfaro, izaskun.lopezsamaniego, gonzalo.zarza}@globant.com
[2] Facultad de Informática, Universidad Complutense de Madrid, Spain
econcepc@ucm.es

Abstract. This article addresses the banking interaction needs of the senior population, devoting special attention to the unique needs of this age demographic delineated by the natural cognitive decline caused by ageing. Despite being a compelling business audience, the financial sector, and particularly banking institutions, have significantly underestimated the importance of designing a customer experience that is better adapted to the needs of senior users. A compelling yet achievable solution to this problem is the adoption of AI-based technologies, such as dynamic user interfaces, biometric authentication methods, automatic recommendations, and experience-based APIs. The curated adoption of these tools will enable banks to better serve this increasingly growing population by creating seamless and respectful experiences for their senior customers and, above all, will provide them greater autonomy.

Keywords: Elderly People · Digital Solutions · Business Model · Social Responsibility · Artificial Intelligence.

1 Introduction

The increasing ageing of the population, in combination with the spread of digitalisation of businesses and services, has introduced a growing urgency to bring to centre stage the needs of elderly people. This statement is not a mere headline, since 33% of Europe's population will be aged 60+ in 2050 [1], [2], compared with 24% in 2016 [3]. Focusing on this age demographic segment as a distinct target constitutes, not only a considerable source of income for the financial sector but, more importantly, a social responsibility [4].

As Srividya highlights in [4], there are nearly 104 million elderly persons just in India, according to the 2011 Census. This implies a huge potential for the financial institutions to expand their customer base and build a reputation for old age-friendly banking service. While banks and other financial institutions do offer products for senior citizens, they do not seem to have developed a comprehensive program to fully address the middle to low income older adults' unique financial concerns [4].

Despite older groups get much higher income and loyalty ratios, it is common for companies to promote themselves directly to younger age demographics, regardless of the fact that they are typically the segment with the least purchasing power, propensity

to pay, as well as loyalty. As Du Toit and Burns point out, "Digital channels often are not as effective as banks would like; many consumers trying these channels contended with various problems and wound up visiting a teller or calling a contact center anyway" [5]. In contrast, older groups are recurrently not prioritized, or directly ignored, in digitisation processes.

In this context, the COVID-19 pandemic has highlighted the increased vulnerability of our aging population, as well as the need to rethink the basis for the interaction design in the current digital business models.

The development of custom-made functionalities for our elderly is not only a need or a social responsibility, but it is also a great business opportunity, as well. The rest of this paper is devoted to the analysis of plausible and economically-viable AI-powered alternatives to overcome this problematic. First, the background and motivation of analysis are presented in Section 2. Then, the proposed solution approach is introduced and described in Section 3. Finally, a brief discussion and conclusions are drawn as closure in Section 4.

2 Background and Motivation

2.1 What happens when our brain starts to fail?

Cognitive deterioration commonly starts in almost all people by the age of 64, and this process increases along with age or the appearance of neurological diseases, as illustrated by Mustafa et al. [6] in Fig. 1. Auspiciously, there are different technologies that can be of great help to solve the problems that can come with this deterioration, such as biometrics, experience-based API Patterns, or automatic recommenders.

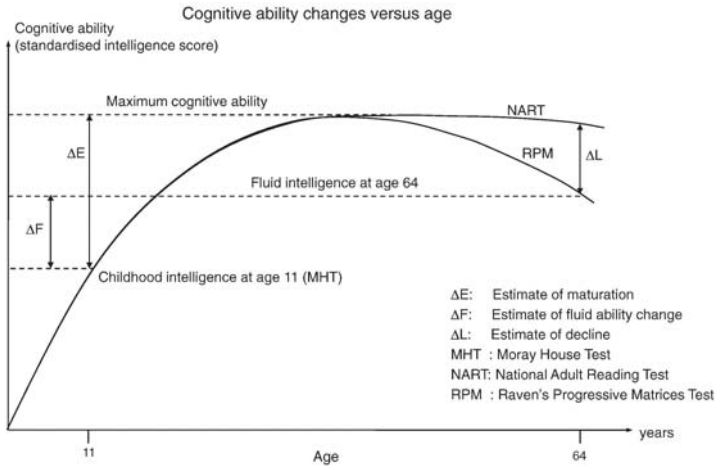


Fig. 1. Cognitive Ability Changes Versus Age (Source: [6]).

2.2 Where should we start?

Undoubtedly, the financial industry, and specially the banking sector, is one of the most important economic and social actors in the world. Over the last years, the reduction in the number of bank branches, along with their corresponding ATMs, has forced bank customers to utilize self-service digital channels, putting elderly customers in a vulnerable situation. The graphs in Figs. 2 and 3 illustrate the magnitude of this paradigm shift, depicting how banks have reduced the number of offices and ATMs in the Eurozone and the United States during the last decade.

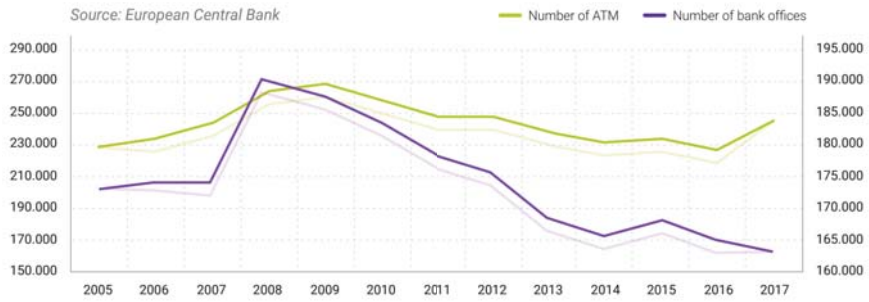


Fig. 2. Bank Branches and ATMs Evolution in the Eurozone.

Fig. 3. Branches Evolution in Europe and USA per 100,000 inhabitants (Source: [7]).

2.3 How can we help?

Technology is a tool that can help simplify our lives. There is relevant evidence that highlights the positive impact of technology as the foundations for creating a new dif-

ferential digital experience for the elderly, with a strong focus on their own autonomy. Nowadays, it is possible to leverage relevant data and scientific evidence to identify significant opportunities to create new respectful and pragmatic multichannel experiences for the senior population. It is now possible to harness state of the art technology to design experiences that require less memory and attention, and truly adapting the experience to the users, and not the other way around.

Some of the technological aspects that we need to take into account to address this problem:

- **Creating a dynamic UI-UX with interactions adapted** in response times, fonts and sizes and only necessary actions and features [8]. By working on accessibility and information architecture, we can reduce the cognitive workload and flatten the learning curve. For example, we can improve accessibility by adapting the font size to increase readability. We can work on information architecture to increase findability and remove those functionalities in the website that could be useless for these customers.
- **Using voice or facial recognition to order or confirm** operative payments, transfers, etc. The elder people will no longer need to memorise their personal access keys. The introduction of facial and voice recognition is a simple and proven way of authentication which eliminates the risk of forgetting a password or introducing erroneous information. These techniques are suitable for various scenarios such as digital and phone banking.
- **Adapting the information, management, and monitoring channels** to the older target audience, understanding exceptions to self-management and helping through remote access. Not everyone is able to operate by means of a mobile application nor understanding the provided information, but all the customers could require assistance or information. Accordingly, we are proposing to revisit the prevailing trend of promoting self-service and developing a digital omnichannel strategy. Although we consider this trend adequate in general terms, we highlight the need for establishing exceptions regarding information access and e2e assistance for the groups we are analysing in this paper.

3 Our proposal: technology to the rescue

In the following paragraphs, we introduce and describe the outcomes of our efforts to empower the senior population in their access to banking services. We have put special emphasis in the study of the applicabilities of three different AI-enabled solution families: automatic recommenders (section 3.1), experience-based APIs (section 3.2), and biometrics (section 3.3).

The widespread use of *automatic recommenders* has the potential to enable a meaningful improvement of the user experience. In the case of our older, senior demographic, the use of recommenders can range from the selection of menu choices to be shown to the user while navigating, to the type of financial products that would interest a senior customer.

Similarly, the concept of *experience-based API* leads to many potential applications. Concerning the older demographic, this approach would help to design a set of customer-centric services with a special focus on senior needs which combined with the use of recommendation techniques, the API can be adequately adapted to serve the needs of the devices and the user profiles.

In addition, the use of *biometrics* could strongly simplify the authentication process. A facial recognition interface, enhanced with the use of voice recognition, would allow older demographics of users a seamless login and easy access to the services she/he needs. The use of special devices, such as totems in a bank office, would also reduce the time spent queuing and give them access to the personal assistant line immediately.

3.1 Automatic Recommenders

Recommender systems are effective tools to predict user preferences and suggest choices in a complex choice-making context [9]. They have become obvious components in applications such as electronic commerce platforms and information query tools, providing suggestions that smartly prune large sets of options so that users are directed toward those that best meet their needs and preferences.

There is a broad range of recommendation techniques such as content-based, collaborative, knowledge-based, and others. They all provide both advantages and disadvantages, and the best choice depends on the context of the problem we need to solve [10].

As we review the existing recommendation techniques along with their main features, it is noteworthy to mention that there are several models of classification. In this paper, we have adopted the Burke perspective [10].

Collaborative recommendation is the most popular strategy, both widely implemented and well proven. The collaborative approach focuses on gathering the ratings provided by the users about the diverse choices and finding commonalities among users according to their rating behaviour. There are several variations of this technique: binary ratings, percentage of affinity ratings, predictors based on Bayesian networks, neural networks, and latent semantic indexing. A valuable feature of the collaborative strategy is its ability to provide recommendations outside-the-box, that is, the ability to find options that the user has never rated, based on the choices of other similar users. The main shortcoming of this approach is the well-known "ramp-up" problem [11]. This term actually refers to two distinct but related problems: the new user (a user with few or no ratings cannot be classified) and the new option (an option with few or no ratings can barely be recommended). Another common problem is data sparsity, which happens when the space of ratings is sparse, that is, when only a few users have rated the same items.

Demographic recommender systems base their operation on categorizing the users according to their personal attributes and make recommendations based on demographic classes. The representation of such demographic information varies greatly from one system to another. The classification criteria depend on the objectives that the system must achieve. The main advantage of this approach is that it does not require a history of user ratings, so it can provide recommendations even with no previous user activity.

A content-based recommendation system stores information about the features of the different options available, and considers the ratings provided by the users to map the user behaviour with the features. Leveraging this information, the recommender can estimate the rating that a user would give to a particular option. Content-based techniques do have a start-up problem in that they must accumulate enough ratings to build a reliable classifier.

Utility-based recommenders generate a utility function that allows them to make suggestions based on a calculus of the utility of each option for the user. Developing such a utility function for each user is the key non-trivial activity for this strategy.

Knowledge-based recommendation analyses the needs and preferences of the users to understand what they need.This kind of system has knowledge about how a particular option meets a particular user need, and can reason about the relationship between a need and a possible recommendation. Such knowledge can be gathered and expressed in multiple ways, depending on the context of operation. Both utility-based and knowledge-based techniques do not have ramp-up or sparsity problems since they do not base their recommendations on accumulated statistical evidence.

Taking into consideration the limitations of single strategies, we have evaluated hybrid strategies to overcome the constraints of single techniques by combining them. This analysis does not intend to be an exhaustive list of all existing hybridisation strategies, but a representative panorama of the most interesting combinations.

In the mixed combination, the recommendations from two different recommenders are presented at the same time to the user [10]:

1. The Cascade approach pipelines the output of one recommender as the input to another one, which refines the recommendations given by the previous step.
2. The Weighted approach, where several recommendation techniques are combined together, according to a set of numerical weights, to produce a single recommendation.

Most commonly, collaborative filtering is combined with other techniques in order to avoid the ramp-up problem. The combination of collaborative filtering with content-based and/or demographic techniques is one of the most frequently used [12], [13], [14].

The navigation across a banking website is one of the first candidates to be enhanced with automatic recommenders since it is not always user-friendly, especially for senior users. The customer is commonly required to navigate through a large number of options with technical names that are not always intuitive. A proper study of the needs and behaviours of senior users would lead to the development of a recommendation-based assistance system to help them during the interaction.

3.2 Experience-based API

An API ecosystem provides secure access to data resources through firewalls, proxies, and API gateways, allowing for new business models and offering a platform that supports the development of new digital products [15].

The concept of Experience-based API is a solution for enriching and/or personalizing the interaction between producers and consumers [16]. Concerning the older demographic, this approach would help to design a set of customer-centric services with a special focus on senior needs.

The most obvious application of the experience-based approach is the tailoring of pre-existing APIs to better meet the needs of the consumers, as depicted in Fig. 4. Another application, closely related to the previous one, is the abstraction of the source. This introduces an additional layer of abstraction that enhances the decoupling of consumers and producers and simplifies the coexistence of different versions of the same API.

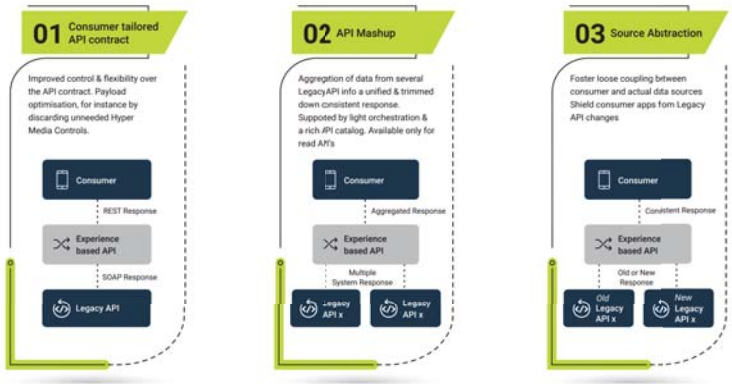


Fig. 4. Experience-based API Patterns (Source: [16]).

Lastly, the mashup pattern allows for building a composite API that provides an integrated view of the resources which can perfectly be backed by different systems/API. This pattern can be very helpful when building tailored views for different consumers of a system of resources.

Combined with the use of recommendation techniques, the API can be adequately adapted to serve the needs of the devices and the user profiles developing a new layer of mediation or abstraction is placed between the consumers of the API and the providers of the API. The published resources may be adapted to mobile, web or IoT scenarios depending on customer and/or security needs.

There are multiple backgrounds in which an Experience-based API approach can be helpful [16]:

- **Adapting to consumer needs:** the available API is coupled to an existing product but the consumers need an adaptation of the exposed resources.
- **Hiding complexity:** in an ecosystem in which there are resources that belong to multiple back-end applications, databases or other sources, providing a homogeneous access to this variety of resources is a must, and the use of this approach allows for abstracting the consumer from complexity.
- **Data governance:** when it is necessary to synchronize data from a variety of services, even when they belong to different business domains. Many departments within an organization make their own purchasing decisions for the products they use and both central control and data governance can be lost.
- **Properly manage application integration expectations:** a digital business application is being introduced to the market and customers of this application will expect integration with the SaaS applications they use within their organization.

3.3 Biometrics

In this paper, we will adopt the simple yet accurate biometrics definition provided by Jain et al. in [17]: "Biometrics is the science of recognizing the identity of a person based on the physical or behavioural attributes of the individual such as face, fingerprints, voice and iris".

There are two common tasks that facial recognition systems perform: verification and identification. Verification refers to the unambiguous identification of a person. Some typical usages of this technique are checking the user's face to unlock smartphones and automatic passenger screening during plane boarding [18]. There are two types of identification based on the awareness and cooperation of the person being identified. If the person that is being identified is aware and tries to facilitate the process, it is "cooperative identification". On the contrary, if the person is unaware that they are being identified, it is "non-cooperative". This variable has an obvious impact on the accuracy of the recognition process.

In the last decade there has been a technological breakthrough in how the use of biometrics can strongly simplify the authentication process. A facial recognition interface, enhanced with the use of voice recognition, would allow older demographics of users a seamless login and easy access to the services she/he needs.

The use of special devices, such as totems in a bank office, would also reduce the time spent queuing and give them access to the personal assistant line immediately. The range of potential applications continues to grow: border control, user authentication, law enforcement, control facilities, etc.

According to Crumpler, "In ideal conditions, facial recognition systems can have near-perfect accuracy. Verification algorithms used to match subjects to clear reference images, like a passport photo, can achieve accuracy scores as high as 99.97% on standard assessments like NIST's Facial Recognition Vendor Test (FRVT)", and highlights that even banks feel comfortable to use these techniques to log their users [18]. However:

- Lighting and positioning play a vital role to achieve a high level of accuracy so the facial features of the individuals must be clear. The error rate can rise from 0.1% under ideal conditions to 9.3% in a real-world scenario [19].
- The awareness and cooperation of the person being identified has an obvious impact on the accuracy of the recognition process.

For this reason, commercial algorithms are often set to only return a match if they have a certain degree of confidence in their assessment to avoid false positives, especially in those scenarios where identifying the wrong person can have severe consequences, such as an ATM authorisation system.

The use of confidence thresholds can significantly lower match rates for algorithms as they force the system to discard positive but low-confidence matches [18]. This authentication method can be combined with other biometric technologies to ensure a completely accurate authentication, such as the application of voice-based authentication.

4 Discussion

Undoubtedly, the introduction of cutting-edge technologies will enable the design and implementation of a holistic experience for senior customers that will significantly improve their banking quality of experience (QoE). However, it is worth noting that despite great advances made thanks to AI-powered technologies and automation techniques, there are still important challenges that need to be overcome in terms of User and Customer Experience. For instance, the great potential of facial recognition can still be ballasted by inadequate lightning conditions or by poor quality datasets.

Notwithstanding, beyond improving the customer experience, the adoption of AI-powered solutions are already making it posible to improve the day-to-day lives of millions of people who are deprived of their autonomy when operating with their banking entities. It is now possible to seriously consider an issue that has been largely ignored: that the digital transformation process must take into account all age demographics in society, and not just the youngest.

One of the lessons that COVID-19 has taught us is that the older age demographic is one of our most vulnerable groups, and not only because of the incidence of the disease.

Fortunately, all the techniques discussed above are facets of a combined architecture exclusively oriented to increase the welfare of senior customers and help them to improve their relationships with their banks.

Bibliography

[1] Rafal Chomik and John Piggott. Population Ageing and Social Security in Asia. *Asian Economic Policy Review*, 10(2):199–222, 2015. https://doi.org/https://doi.org/10.1111/aepr.12098. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/aepr.12098.

[2] J. P. Michel, B. L. Beattie, F. C. Martin, and J. D. Walston. *Oxford Textbook of Geriatric Medicine*. Oxford Textbook. Oxford University Press, 2018. ISBN 9780198701590. URL https://oxfordmedicine.com/view/10.1093/med/9780198701590.001.0001/med-9780198701590.

[3] United Nations. *World Population Ageing 2017 Highlights*. United Nations Documents. UN, 2019. ISBN 9789211515510. URL https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf.

[4] P. N. Srividya. Making banking a delight for the elderly, May 2018. URL https://www.edgeverve.com/blogs/making-banking-delight-elderly/. [online; accessed on April 2022].

[5] G. Du Toit and M. Burns. Customer Loyalty in Retail Banking: Global Edition 2016, November 2016. URL https://www.bain.com/insights/customer-loyalty-in-retail-banking-2016/. [online; accessed on April 2022].

[6] Nazahah Mustafa, Trevor S. Ahearn, Gordon D. Waiter, Alison D. Murray, Lawrence J. Whalley, and Roger T. Staff. Brain structural complexity and life course cognitive change. *NeuroImage*, 61(3):694–701, 2012. ISSN 1053-8119. https://doi.org/https://doi.org/10.1016/j.neuroimage.2012.03.088. URL https://www.sciencedirect.com/science/article/pii/S105381191200376X.

[7] José Francisco López. ¿Desaparecerán las sucursales bancarias?, May 2018. URL https://economipedia.com/actual/desapareceran-las-sucursales-bancarias.html. [online; accessed on April 2022].

[8] W3C. Developing Websites for Older People:How Web Content Accessibility Guidelines (WCAG) 2.0 Applies, January 2018. URL https://www.w3.org/WAI/older-users/developing/. [online; accessed on April 2022].

[9] Robin Burke. Knowledge-based recommender systems. *Encyclopedia of library and information systems*, 69(Supplement 32):175–186, 2000.

[10] Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[11] Joseph A. Konstan, John Riedl, Al Borchers, and Jonathan L. Herlocker. Recommender systems: A grouplens perspective. In *Recommender Systems: Papers*

from the 1998 Workshop (AAAI Technical Report WS-98-08), pages 60–64. AAAI Press Menlo Park, 1998.

[12] Mark Claypool, Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes, and Matthew Sartin. Combining content-based and collaborative filters in an online newspaper. In *Proceedings of ACM SIGIR workshop on recommender systems*, volume 60, pages 1853–1870. Citeseer, 1999.

[13] Michael J. Pazzani. A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13(5):393–408, 1999.

[14] Barry Smyth and Paul Cotter. A personalised TV listings service for the digital TV age. *Knowledge-Based Systems*, 13(2-3):53–59, 2000.

[15] Daniel Jacobson, Greg Brail, and Dan Woods. *APIs: A strategy guide*. O'Reilly Media, Inc., 2012.

[16] Ross Garret. API Mediation: Why You Need an API Experience Layer, May 2017. URL https://nordicapis.com/api-mediation-why-you-need-api-experience-layer/. [online; accessed on April 2022].

[17] Anil K. Jain, Patrick Flynn, and Arun A. Ross. *Handbook of biometrics*. Springer Science & Business Media, 2007.

[18] William Crumpler. How Accurate are Facial Recognition Systems–and Why Does It Matter?, April 2020. URL https://www.csis.org/blogs/technology-policy-blog/how-accurate-are-facial-recognition-systems-âĂŞ-and-why-does-it-matter. [online; accessed on April 2022].

[19] National Institute of Standards and Technology, U.S. Department of Commerce. FRVT 1:1 Leaderboard, February 2020. URL https://pages.nist.gov/frvt/html/frvt11.html#overview. [online; accessed on April 2022].

Process Optimization in the Steel Industry using Machine Learning adopting an Artificial Intelligence Low Code Platform

Federico Walas Mateo[1]; Andrés Redchuk[2]
[1] *Universidad Nacional Arturo Jauretche, Buenos Aires, Argentina,*
Universidad Nacional de La Plata, Facultad de Informática.
fedewalas@gmail.com
[2] *ETSII. Universidad Rey Juan Carlos. Madrid. España,*
Universidad Nacional de Lomas de Zamora. Facultad de Ingeniería. Argentina.
andres.redchuk@urjc.es

Abstract.

Traditional industries like steelmaking, are in the spotlight for the need of improving processes towards net zero emissions. This article presents a case on a new business model to ease the adoption of Machine Learning (ML) to optimize industrial processes, applied to a blast furnace at a steel company. The focus of the paper is to illustrate the way a ML platform with a Low Code solution approach can give results in two months to optimize a production process at a steel mill. The methodology used in the case allows obtaining a data model to be validated in less time than conventional approaches. This work pretends to give more light to the use of industrial data and the way traditional industries can evolve towards the industry 4.0 paradigm. The adoption of the low code solution is based on lean startup methodology. The cycle to obtain valid results includes the involvement of people from the process as well as analytics experts. At the end it can be seen that the solution contribute to improve Operational Equipment Effectiveness (OEE) and lower energy consumption. Besides process operators became empowered with the predictions that give the platform.

Keywords: Lean Startup methodology, OEE, AI/ML, Low code platform, Net zero emissions, Industry 4.0.

1. Introduction

This paper aims to go deeper from previous research [1] and addresses the adoption case of a Machine Learning (ML) platform through a new business model [2] in the steel industry to add value in a critical process, and evolve towards a smart manufacturing or Industry 4.0 (I4.0) model. The case pretends to highlight the opportunities that low code Industrial platforms, this time an Artificial Intelligence (AI) solution is

featured, can generate to transform traditional factories toward a smarter and more efficient environment, and how people on the shop floor can be empowered by the information these types of solutions produce.

Steel making is a key industry in terms of environmental sustainability in the Net Zero Carbon emissions objective. Going further in the level of criticality that steel industry has reached, [3] points that there is an unprecedented focus on safety, environment impact and sustainability. According to this paper there is a growing concern in maintaining the social license to operate. All those challenges, combined with ever-changing commodity prices and demand paradigm, have accelerated the steel industry's pursuit of sustainable solutions that are capable of helping improve operational performance across the entire value chain.

The previous work [1] highlight the disruptive potential of AI, and ML, to generate new business models and opportunities for startups according to [5]. It is remarked the rise to radical new operating models, this paper pretends to illustrate a case linked to this affirmation, and how it can be applied in the industrial environment.

An issue to consider when talking about I4.0 model is the intense use of data to leverage decision making and facilitate management, according to [6] there is an important quantity of legacy, enterprise, and operational systems data that is not properly used. The article remark that many industries are sitting on a goldmine of unexplored historical, legacy, and operational data from their manufacturing execution systems (MESs) and enterprise resource planning systems (ERPs), among other software sources, and they cannot afford to miss out on its unexplored potential. However, only 20–30% of the value from such available data-at-rest is currently accrued.

Finally this paper approaches lines raised from previous works [7] regarding the implications for people in the process and digitalization, for the success of AI/ML as process optimization driver under I4.0 paradigm. The methodology presented in this work intends to ease the the adoption of the new model through a co creation process between experts from the industrial process, and software engineers from the AI platform.

2. Methodology. A Lean startup approach proposed by an AI Low code Platform

The methodology to be applied in the case considers the opportunities to adopt AI/ML in the steel industry. The approach is based on Low-Code solution [8], and Lean Startup methodology [9] to achieve results in less time than conventional adoption processes of analytics solutions, making possible to democratize artificial intelligence and machine learning in traditional industrial environments.

The value proposal of the new business model presented by the Low Code Platform (LCP) focuses on shortening the adoption cycle of AI/ML in industrial environments by using prebuilt templates for manufacturing processes.

[10] Affirms that LCPs facilitate achieving objectives at the core of business information systems research such as increasing productivity and reducing costs of developing and maintaining enterprise software systems and improving organizations' ability to adapt software systems to rapidly changing requirements, and empowering users. The author also points that it is not at all clear what distinguishes LCPs from existing

software development facilities, such as classical integrated development environments (IDEs)and tools for model-driven development (MDD).

The LCP used in the case of this paper, offers seven templates: Forecasting, Anomaly Detection, Optimization, Simulation, Failure Prediction, and Defective Part Prediction. The template to use in each case depends on the nature of the process and the opportunity to address.

These preconfigured frameworks allow users in the shop floor to understand easily what I aimed with the data model, have it earlier and evaluate results faster. The adoption cycle begins working with the model offline, in a co creation space between process and software sides. Once created the analytic model is fed by data from the historian database. This way, the operators evaluate results and gain confidence by iterating through multiple experiments.

Finally, the model is deployed into operations, when is fed with real-time data from the industrial processes. The solution is offered in as a service (SaaS) model, ingesting data in a data series format from MES; or Industrial Internet of Things (IIoT) platforms. This way the AI/ML software is integrated in the industrial operation with the shop floor solution, and work with data in real time to generate the predictions to the process operators. A complex issue for the industrial environment, and a weakness reported by the study presented by [11] is cybersecurity. This issue is tackled by hosting the platform at Azure Microsoft Cloud (https://azure.microsoft.com/). This cloud solution also provides infrastructure needed for the requirement of AI/ML models.

To validate the effectiveness of the platform and have results in a shorter period than traditional approaches, a Lean Startup methodology is used. This way non-value adding activities are minimized and people from the industrial process can be involved earlier, while introducing the new solution. The Lean Startup methodology has three key principles: to replace planning with experimentation, the ‘getting out of the building’ approach and lastly, agile development [12].

The experimentation process is described by the Build-Measure-Learn feedback loop consisting of three steps: build, measure, and learn. In the first step, build, it is essential to create a Minimum Viable Product (MVP) using as less resources as possible after identifying the most important hypotheses [9]. The goal of building an MVP is to identify the proposed solution’s potential [13] and the value for the customer. The measure step aims at collecting data that can verify or dismiss the hypothesis made about the solution to be offered [9]. In the learn step, the goal is to know about the hypotheses from collected data. The learning process shows if an underlying hypothesis can be verified indicating that the MVP matches a customer requirement.

3. Adoption Case in the Steel Industry

The case examined was the process in the blast furnace of a North American steel company, where there was a need to optimize the production line by improving the quality and quantity of performance of the BF. As mentioned before, these are complex and dynamic processes that generate hundreds of data types with high variability. This leads to management challenges to maintain consistency of production quality, lower energy consumption, and use less silicon.

It should be considered that inconsistencies in quality in the foundry process generate excess scrap, rework, and production delays, among other issues. Traditional quality control was done through samples that were analysed in a laboratory. The problem with this methodology is that if there was an inconsistency, the problem was solved in a reactive way, when it had already occurred.

On the other hand, it must be considered that the process ran in three eight-hour shifts, and the operators were different, each with their own criteria for and experience in the process. Then, its control varied depending on each person. In addition, the smelting process included a large number of chemical and thermodynamic variables that were very difficult to monitor and control.

The firm generated real time data from the production processes through a MES software, OSI PI from OSIsoft [14]. This data is stored in historian-type databases.

4.1 Adoption process and Results

To address the improvements in the process, the steel company chose the solution proposed by the AI platform with the following objectives:

1. Identify the parameters that influence the maximization of production performance using process data generated by the OSI PI platform.
2. Predict product quality and performance at different intervals, and use those predictions to adjust control parameters in real time, thus maximizing product quality and performance.
3. Apply a proactive policy for the resolution of problems and anomalies of the process based on having advance information about what happens in the casting process.
4. Institutionalize the best practices of the operators and provide those responsible for the information processes that facilitate decision making.

The case was developed using the Build-Measure-Learn feedback loop detailed previously in section 3 and displayed in figure 1 at the bottom of this section. The first phase of the methodology, BUILD, started after defining the adoption scope and its objectives. The process experts at the steel works self-trained with the support of experts in data analysis at the Startup. Meanwhile, both parties collaborated with to clarify doubts and specific topics that arises at the data experts' side, about metallurgy issues and characteristic of the process at the BF.

An issue to highlight that make possible the adoption of the LCP, is that there was enough data available generated by the MES solution, it only needed to be normalized. Then the process engineers from the steel works began working with the analytics and math experts to analyze and understand the methodology and technical details of the milling process. The first step of the methodology took three weeks.

An observation that should be done is that there is no existing baseline for this use case, the operator could not predict such behavior by the traditional operation. It only can be estimated based on a quick metallurgical balance. In this case, the operation depend on the intuition of the operator and even on management policies.

The following stage according to the Lean Startup methodology was the MEASURE one. In this phase a model was produced in the software platform and the plant users evaluated the use case with historical data ingested on the platform.

A data exploration and evaluation of data model by the BF Process Managers was performed and analytics experts from the software startup provided support at all times. The process engineers went thought the model, and finally they validated it.

The second phase of the methodology was completed with the model going live at the platform on line. Real time data from the MES solution started to be ingested to the cloud based solution. The integration was made by software specialists from the Startup. Besides the software solution has an Aplication Process Interface (API) that simplifies the integration with OSI PI platform. This way the process users accessed to the ML solution with continuous support from the provider. This cycle took six weeks.

The last stage was the one called LEARN, according to Lean Startup methodology. In this last phase the plant staff recognized the value that the use case contributed. Results were analyzed by the software supplier and plant engineers. The use case was reviewed on a platform mapping session where both teams gained deeper insights into data, business value, and complexity of previously identified use cases.

The adoption process took two months until project kick-off in order to have the model on line ingested with real-time data, and generate predictions to suggest better operation conditions to the furnace operators. In this period, the people in the process were taught how to use and understand the information given by the solution, while people from the platform provider adjusted the model. The validation was done together by the software engineers and plant operators.

The platform made it easy to define the relationships between variables that significantly affected the production result, such as quantity of silicon added and fuel consumption. The AI models collaborated to determine which of the control variables had a positive or negative impact on production quality. By applying ML-reinforced learning through the forecasting template.

As a result of predictions generated by the ML LCP, the steel firm obtained a competitive advantage by improving OEE, and achieving consistency in quality with silicon content always at a normal range. Additionally, by defining set points in real time for fuel consumption, energy savings were facilitated. The OEE improvement was achieved mostly by reducing the percentage of scrap, thus reducing costs and producing higher quality steel. There was a triple optimization for the firm: reducing costs, improving customer services revenue. Moreover by improving energy consumption the firm is aligned with Net Zero Carbon emissions objective.

The use of ML methods employed alongside domain knowledge of BF specialists facilitated a greater knowledge to process people that was augmented continuously. This allow to work on continuous improvement and optimize Key Performance Indicators (KPIs), evolving toward more mature management systems.

4. Conclusions

The case addressed the adoption of a new business model based on AI/ML in a traditional industrial environment and on the complex steel smelting process in a blast furnace. The solution came from a startup, and had an advantage over new methodologies

that shortened times and simplified adoption of analytics in the industry. The distinguishing feature of the approach with respect to other works was the Low-Code approach, which eased the use of AI/ML by process operators with little math or statistics knowledge, and shortened adoption times radically.

In line with the above paragraph, change management was simplified toward an agile approach; the new business model generated results faster and with lower complexity than traditional analytics solutions. This was because the model proposed by the platform of the case, is centred on the end user, the process operator. The Low-Code model from the startup eased time-consuming and risky tasks linked with software development and algorithms. In this way, the business model proposed in this paper helped to democratize AI/ML in traditional industries, making the adoption of these types of tools easier for a broad segment of people.

The improved process was based on predictive analytics aimed at exploiting the huge treasure of legacy operational data and overcoming some challenges of real-time data analytics. The potential of the proposed approach is high in traditional industries that have not benefited from the advancements of Industry 4.0, and in most cases have just started investigating the potential of data analytics and machine learning for the optimization of their production processes.

On the other hand, the need for a large amount of operational data is an important limitation for various traditional industries that are still lagging on digitalization. This issue is an important weakness of industries that could use data models to improve the remaining useful life (RUL) of heavy equipment, reduce carbon footprints, and reduce non-value-adding activities, among other wasted opportunities.

Another point to highlight is the opportunities that the smart manufacturing paradigm opens to startups and their innovative business models that offer solutions to ease the adoption of AI/ML in the industry. This case showed how a business model using AI/ML focused on a No-Code/Low-Code strategy could shorten implementation cycles from several months to a couple of months, as shown in the case at the steel firm.

Using information captured from the process that is generated by shop floor integrated platforms such as MES or IIoT, seems to be a key point to improve processes, and evolve towards AI/ML models. Not doing so should be considered a waste that translate into lack of competitiveness for industrial firms.

Finally it could be seen that I4.0 model ease data into knowledge to empower people at the industrial shop floor. From the technological side, the case show how OT and IT is integrated from sensors at the plant, MES platform, LCP, and cloud infrastructure to provide knowledge to improve the productive process.

5. Acknowledgements

The authors wish to thank two research projects that made possible the content of this article. The projects are “Metodologías de abordaje al modelo Industria 4.0 en PyMEs, el rol de la Empresas de Base Tecnológica (EBT), los recursos humanos, y el ecosistema de Innovación”, Resolution (R) No. 183-21-UNAJ INVESTIGA Program; and

"Mejora de Procesos, Optimización y Data Analytics: Aplicación en Procesos e Industrias de Interés Regional Mediante Estudios de Casos Reales", approved by a resolution of the UNLZ Engineering Faculty.

References

1. Redchuk, A.; Walas Mateo, F. (2022). New Business Models on Artificial Intelligence—The Case of the Optimization of a Blast Furnace in the Steel Industry by a Machine Learning Solution. Appl. Syst. Innov. 2022, 5, 6. https://doi.org/10.3390/asi5010006A.
2. Osterwalder. Y. Pigneur and C.L. Tucci (2005). Clarifying Business Models: Origins, Present, and Future of the Concept. Commun. Assoc. Inf. Syst. 2005, 16, 1–40.
3. A Toktarova, V. Walter, L. Goransson, and F. Johnsson, (2022). Interaction between electrified steel production and the north European electricity system. Applied Energy. Volume 310, 15 March 2022, 118584
4. A. Endl, M. Tostb, M. Hitchc, P. Moserb and S. Feielb (2019) Europe's mining innovation trends and their contribution to the sustainable development goals: Blind spots and strong points. Resources Policy, https://doi.org/10.1016/j.resourpol.2019.101440
5. W. Pfau and P. Rimpp (2021). AI-Enhanced Business Models. In Digital Entrepreneurship, Digital Entrepreneurship, Future of Business and Finance; Springer International Publishing: Berlin/Heidelberg, Germany, 2021. https://doi.org/10.1007/978-3-030-53914-6_7.
6. K. Lepenioti, M. Pertselakis, A. Bousdekis, A.L Fenareti Lampathaki, D. Apostolou, G. Mentzas and S. Anastasiou (2020). Machine Learning for Predictive and Prescriptive Analytics of Operational Data in Smart Manufacturing. In Advanced Information Systems Engineering Workshops Proceedings of the International Conference on Advanced Information Systems Engineering, Melbourne, CAiSE 2020, VIC, Australia, 8–12 June 2020.
7. F. Walas Mateo; A. Redchuck. (2021). IIoT/IoT and Artificial Intelligence/Machine Learning as a Process Optimization Driver under Industry 4.0 Model. Journal of Computer Science & Technology, vol. 21, no. 2, pp. 170 -176. doi.org/10.24215/16666038.21.
8. J. Cabot (2020). Positioning of the low-code movement within the field of model-driven engineering. In Proceedings of the MODELS'20: ACM/IEEE 23rd International Conference on Model Driven Engineering Languages and Systems, Virtual Event, Canada, 18–23
9. E. Ries (2011). The Lean Startup. New York: Crown Business.
10. A. C. Bock, U. Frank. (2021). Low-Code Platform, Bus Inf Syst Eng 63(6):733–740
11. T.H. Davenport (2020). Is AI Getting Easier? Forbes Media LLC.: New York, NY, USA, 2020.
12. S. Blank (2013). Why the Lean Startup changes everything. Harvard Business Review, May 2013.
13. Kerr, W. R., Nanda, R., Rhodes-Krop, M. (2014). Journal of Economic Perspectives. Volume 28, Number 3, 2014. Pages 25–48. DOI: 10.1257/jep.28.3.25
14. OSIsoft Website. 2021. Available online: www.osisoft.com (accessed on 26 July 2021).

Improving audio of emergency calls in Spanish performed to the ECU 911 through filters for ASR technology

Marcos Orellana[1 [0000−0002−3671−9362]], Angel Alberto Jimenez Sarango[1 [0000−0003−0018−4535]] and Jorge Luis Zambrano-Martinez[1 [0000−0002−5339−7860]]

Universidad del Azuay, Cuenca, Ecuador
{marore, jorge.zambranno}@uazuay.edu.ec, ajimenez@es.uazuay.edu.ec
https://lidi.uazuay.edu.ec

Abstract. In recent years, Automatic Speech Recognition (ASR) services have performed notable progress in the research efforts of big companies such as Google and Amazon. However, the ASRs are still sensitive to the audio processing quality in other languages. To solve this issue, various speech enhancement algorithms that are the most prominent in improving speech intelligibility were proposed, such as Singular Value Decomposition (SVD), log Minimum Mean Square Error (log-MMSE) and Wiener. By preprocessing the audio files with these algorithms, we seek to reduce the Word Error Rate (WER), which compares the transcription performed by the ASR against a manual transcription. Thus, we can determine the percentage of error that the ASR service has acquired. Results demonstrated that Google is more efficient than Amazon and Vosk counterparts. Also, we decided that applying a Low-pass filter combined with a log-MMSE algorithm to the audio files can substantially reduce the WER percentage of transcription depending on the noise characteristics contained in the audio.

Keywords: Automatic Speech Recognition · Word Error Rate · speech enhancement algorithms · audio quality improvement.

1 Introduction

Automatic Speech Recognition (ASR) has advanced rapidly in the last few years due to continuous improvements. These systems' quality is affected by features used to record the audios and how they are processed by these algorithms, thus reducing their efficiency in speech recognition [14]. For the correct interpretation of the content of audio files in the ASR algorithms, it is necessary to improve the audio quality through processing techniques [17]. Speech enhancement algorithms such as Singular Value Failure (SVD), Wiener or the Minimum Mean Square Error (MMSE) are the most prominent algorithms due to their performance in improving speech intelligibility [9]. Likewise, other techniques seek to improve the audio quality, such as the low-pass, band-pass or high-pass filter,

which enhances the audio quality in real-time while being more efficient than other audio improvement algorithms [16]. There is a wide variety of research about comparing the performance of ASR algorithms by applying the Word Error Rate (WER) index, a measure commonly used to evaluate those algorithms [7],[8], [15]. Also, other researchers have compared the different speech enhancement algorithms [2], [4], [5], but so far, there is no research about ASR services and applying speech enhancement algorithms. Our proposal focuses on preprocessing audio file datasets of emergency calls provided by the Integrated Security Service ECU 911 for using ASR algorithms after the results are compared with transcriptions performed by humans.

2 Related Works

In a speech-to-text conversion, over 73,7% use the WER metric as an evaluation method for the ASR voice recognition [1], and the ASRs are used in some areas such as telephony, military and client services [6]. However, Kepuska & Bohouta [7] demonstrate that Google API is more efficient than open-source APIs such as Microsoft Speech API or Sphinx-4. Authors calculate the WER index by processing English-language audio from different sources. Several studies compare the performance of ASR services as Vascones et al. [15] demonstrate that Amazon Transcribe has a lower average WER percentage of the transcriptions of audio files without any speech enhancement algorithms from the Integrated Security Service ECU 911 than its competitor Google Cloud Speech-to-Text. In another study, Plaza et al. [12] implement a recognition voice model in Spanish with an ASR offline called CMUSphinx.

Nonetheless, Nian et al. [11] conclude that eliminating the noise in an audio file through background noise removal preprocessing helps decrease the WER index by up to 22.1%. In the same way, Shrawankar & Thakare [13] conclude that the traditional Wiener and log Minimum Mean Square Error (log-MMSE) algorithms are frequently used in speech accuracy tests. However, Chen et al. [3] propose a SVD algorithm to remove noise from audio files. Modhave et al. [10] demonstrate that the Wiener algorithm improves speech quality because this algorithm greatly helps to estimate the noise signal in audio. And Meiniar et al. [10] conclude that it is possible to filter the human speech using a band-pass filter and lose very little speech in the audio files analysed. Thus, we propose preprocessing the audio files pro- vided by the Integrated Security Service ECU 911 through the SVD, log-MMSE, Wiener, and low-pass algorithms that have been demonstrated in the research as the best algorithms for improving speech quality and then transcribing them with a low index-WER.

3 Methodology

This section describes the procedure followed to improve the speech quality audio files in four activities, detailed below.

The first step required creating a Google Cloud account to use the Google Speech-to-Text service. Likewise, an Amazon Web Services account must be created to use the Amazon Transcribe service. Both platforms were implemented with a necessary internal configuration, such as creating workspaces for transcriptions and storage spaces called Buckets on the respective media.

For audio preprocessing, an algorithm was created with Python programming language in its version 3.9.9, including all the speech improvement algorithms selected after a rigorous analysis of the previously reviewed scientific papers. This way, the speech enhancement algorithms SVD, Wiener and log-MMSE were selected. On the other hand, we analysed that using a low-pass filter can help reduce the background noise from audio files, and the processing was performed with this filter.

In this methodology step, a script was created with the necessary configurations to implement the chosen transcription services: Google Cloud Speech-to-Text, Amazon Transcribe and Vosk. In addition, cloud storage services called buckets and configurations were used in the cloud service consoles. Moreover, the script for using Vosk's offline service was developed.

One of the metrics to measure the level of transcripts in an ASR system is the WER index. This rate takes a reference transcript that contains no apparent error because it is done manually. Subsequently, the automatic transcriptions performed by Google, Amazon and Vosk were taken as hypotheses. Therefore, the WER index compares both texts considering the words that exist in the reference transcript (N), which had been inserted (I), deleted (D), substituted (S), and presented as a result, besides the percentage error of the transcriptions concerning the reference text. The WER percentage is shown with the following equation: $WER = \left(\frac{I+D+S}{N}\right) \cdot 100$.

4 Experimental Results

As explained in the previous section, the standard metric used to compare the accuracy of the transcripts produced by the ASRs systems is obtained through the WER equation.

In Figure 1, we can observe the results obtained with the WER index of the transcripts of the ASR systems from a set of audio files provided by the Integrated Security Service ECU 911, jointly with the employment of the audio quality improvement filters. These audio files contain sensitive situations, and they are under personal data protection law. In our experiments, Google Cloud performed better using original audio files, reducing the WER percentage. However, the combination of speech enhancement algorithms did not substantially reduce the WER percentage because the audio quality and speech intelligibility were not equivalent. For this reason, the applied algorithms did not distinguish between background noise and the different natural distortions of the voice. Those algorithms tended to eliminate them equally, reducing speech intelligibility in the audio files studied.

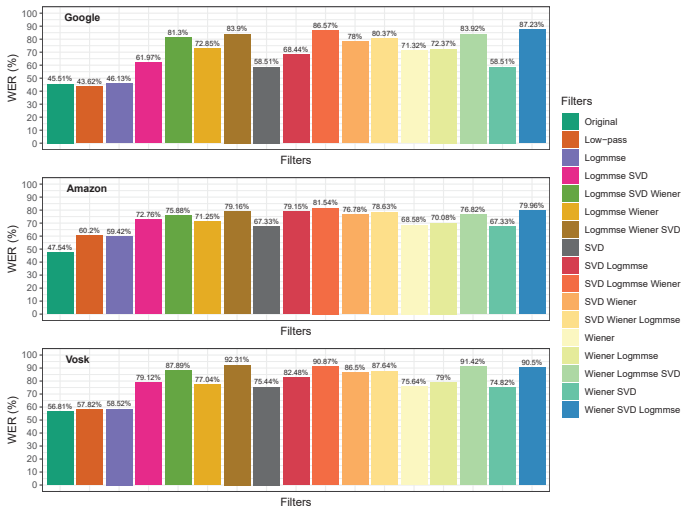


Fig. 1. WER percentage of audio transcripts from ASR systems.

When analysing the results of each audio file transcribed, we observed that a transformed audio file presented a lower WER percentage than the transcription of the original audio file. Other audio files had their WER percentage higher. Figure 2 shows WER percentages of the audio transcripts with the lowest percentages. In comparison, Figure 2 shows the WER of the audio that has the highest rates.

These audio files tested have characteristics that determine the variation of WER percentage. For this reason, audio files with a duration of over three minutes, with a calm conversation between the operator and the caller in an environment with no background noise and with a vocalisation of all the words adequately without the use of a particular lexicon in the Spanish, can identify a significant percentage of words in the audio file. Preprocessing the audio files with the speech enhancement algorithms before applying them to the ASR system as SVD and log-MMSE, they estimated that the noise in the audio signal was not much. The SVD algorithm eliminated part of the speech of an audio file when eliminating the noise from the wave. Although the log-MMSE algorithm estimated signal to noise and the low-pass filter only cleaned the frequencies, the WER percentage was not significantly increased. However, audio files with less than one minute with a conversation without vocalisation adequate, with background noise, and with a lot of regional lexica induced the ASRs to be unable to identify the words optimally and increased the WER percentage. However, log-MMSE and SVD algorithms calculated the excessive noise in the original audio files and removed the noise. Those algorithms did not eliminate the speech contained in the audio file. Applying the low-pass filter removed only frequencies other than the human voice. Hence, the WER index was significantly reduced.

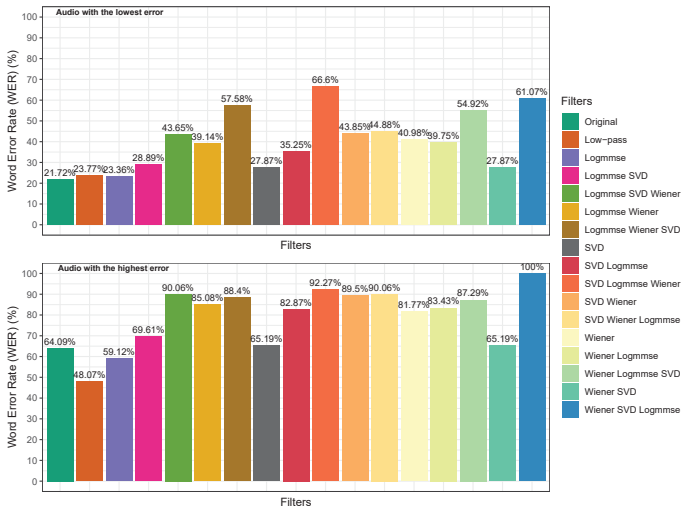


Fig. 2. WER percentage of the audio with the lowest and the highest error.

5 Conclusions

Our research compares transcription quality between the ASR systems offered by Google, Amazon and Vosk after processing the audio files using speech enhancement algorithms such as SVD, log-MMSE, and Wiener and the application of a low-pass filter. The results demonstrate that using ASR offered by Google has a better performance both in the original audio files and in the audios processed by the enhancement algorithms. The WER percentage is reduced depending on the characteristics of the audio files tested, such as the level of background noise, use of a particular lexicon in Spanish, and the length of conversation in the audio files. Using the result obtained as a basis for future work, we plan to develop a classification process for the audio files depending on their characteristics that can be known in advance if one of these filters should be applied, thus considerably improving the transcription of the audio file.

Acknowledgement

This work was partially supported by the vice rectorate of Research at Universidad del Azuay for their financial and academic support, as well as the entire staff in the Computer Science department and the Laboratorio de Investigación y Desarrollo en Informática - LIDI.

References

1. Bhardwaj, V., Ben Othman, M.T., Kukreja, V., Belkhier, Y., Bajaj, M., Goud, B.S., Rehman, A.U., Shafiq, M., Hamam, H.: Automatic speech recognition (asr)

systems for children: A systematic literature review. Applied Sciences **12**(9), 4419 (2022)
2. Chaudhari, A., Dhonde, S.: A review on speech enhancement techniques. In: 2015 International Conference on Pervasive Computing (ICPC). pp. 1–3. IEEE (2015)
3. Chen, X., Litvinov, Y.A., Wang, M., Wang, Q., Zhang, Y.: Denoising scheme based on singular-value decomposition for one-dimensional spectra and its application in precision storage-ring mass spectrometry. Physical Review E **99**(6), 063320 (2019)
4. Dash, T.K., Solanki, S.S.: Comparative study of speech enhancement algorithms and their effect on speech intelligibility. In: 2017 2nd International conference on communication and electronics systems (ICCES). pp. 270–276. IEEE (2017)
5. Gael, P., Chandra, M., Saxena, P., Gupta, V.K.: Comparative analysis of speech enhancement methods. In: 2013 Tenth International Conference on Wireless and Optical Communications Networks (WOCN). pp. 1–5. IEEE (2013)
6. Jamal, N., Shanta, S., Mahmud, F., Sha'abani, M.: Automatic speech recognition (asr) based approach for speech therapy of aphasic patients: A review. In: AIP Conference Proceedings. vol. 1883, p. 020028. AIP Publishing LLC (2017)
7. Këpuska, V., Bohouta, G.: Comparing speech recognition systems (microsoft api, google api and cmu sphinx). Int. J. Eng. Res. Appl **7**(03), 20–24 (2017)
8. Kimura, T., Nose, T., Hirooka, S., Chiba, Y., Ito, A.: Comparison of speech recognition performance between kaldi and google cloud speech api. In: International Conference on Intelligent Information Hiding and Multimedia Signal Processing. pp. 109–115. Springer (2018)
9. Loizou, P.C.: Speech enhancement: theory and practice. CRC press (2007)
10. Modhave, N., Karuna, Y., Tonde, S.: Design of multichannel wiener filter for speech enhancement in hearing aids and noise reduction technique. In: 2016 Online International Conference on Green Engineering and Technologies (IC-GET). pp. 1–4. IEEE (2016)
11. Nian, Z., Tu, Y.H., Du, J., Lee, C.H.: A progressive learning approach to adaptive noise and speech estimation for speech enhancement and noisy speech recognition. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 6913–6917. IEEE (2021)
12. Plaza Salto, J.G., Cristina, S.Z., Acosta Urigüen, M.I., Orellana Cordero, M.P., Cedillo Orellana, I.P., Zambrano-Martinez, J.L.: Speech recognition based on spanish accent acoustic model. Enfoque UTE **13**(3), 45–57 (07 2022)
13. Shrawankar, U., Thakare, V.: Noise estimation and noise removal techniques for speech recognition in adverse environment. In: International Conference on Intelligent Information Processing. pp. 336–342. Springer (2010)
14. Singh, N., Agrawal, A., Khan, R.A.: A critical review on automatic speaker recognition. Science Journal of Circuits, Systems and Signal Processing **4**(2), 14–17 (2015)
15. Vásconez, J.J.P., Ortiz, C.A.N., Cordero, M.P.O., León, P.A.P., Orellana, P.C.: Evaluación del reconocimiento de voz entre los servicios de google y amazon aplicado al sistema integrado de seguridad ecu 911. Revista Tecnológica-ESPOL **33**(2), 147–158 (2021)
16. Wan, F., Yuan, Z., Ravelo, B., Ge, J., Rahajandraibe, W.: Low-pass ngd voice signal sensoring with passive circuit. IEEE Sensors Journal **20**(12), 6762–6775 (2020)
17. Xu, X., Flynn, R., Russell, M.: Speech intelligibility and quality: A comparative study of speech enhancement algorithms. In: 2017 28th Irish Signals and Systems Conference (ISSC). pp. 1–6. IEEE (2017)

Mobile and Web Computing

A *Spotify*-based system for promoting the wellbeing

P. Álvarez[0000−0002−6584−7259], J. García de Quirós, and S. Baldassarri[0000−0002−9315−6391]

Computer Science and Systems Engineering Department
María de Luna, 1 , Ada Byron Building, Zaragoza
University of Zaragoza, Spain
alvaper@unizar.es, jgarciaqg@unizar.es, sandra@unizar.es

Abstract. Different mobile applications and smart systems are being developed to increase users' wellness and happiness. Unfortunately, some of the most recent technological advances in the field of affective computing, Internet of things or service computing have not yet been included in these solutions. In this paper, we briefly present a smart system that analyses the user's emotions during her/his diary activities and configures mood regulation experiences when she/he comes back at home. These emotion-aware experiences are based on the *Spotify* music services and are personalised for each particular user considering her/his musical tastes and preferences. Besides, the system integrates an emotion recognition system based on wearables and artificial intelligence techniques. The recognised emotions are then used to determine the user's mood and to make decisions on the music interventions to be carried out.

Keywords: Smart systems · mood regulation · mobile applications · music · wellbeing

1 Introduction

Positive Computing is a new technological paradigm whose aim is to promote the development of applications for increasing individuals' wellness and happiness [3]. These applications require to recognise the user's emotions in order to recommend her/him activities or media contents that produce a positive effect on her/his mood. Besides, these recommendations should also take in account the user's context and her/his tastes and preferences with the purpose of providing personalised experiences. These requirements imply to integrate the latest advances in the field of affective computing, Internet of things, ambient intelligence and cloud computing.

Many mobile applications and smart environments for improving the wellbeing have been programmed during the last years. The applications are aimed to encourage the user to make physical activities and to take care her/his mental health [2, 9]. Unfortunately, these application use self-assessment methods or human activity recognition techniques to determine how the user feels, instead

of integrating emotion recognition techniques based on wearable technologies. Besides, the interventions proposed for improving the user's wellbeing mostly consist of physical activities and motivating messages. Nevertheless, some example of application that recommends multimedia contents can be exceptionally found [8, 4]. On the other hand, the smart environments are aimed to reduce the user's stress or to regulate her/his mood through experiences that take place in a limited and controlled space, for example, inside a room [5, 10]. These experiences consist of configuring the space conditions for creating a positive atmosphere, for example, through the control of the lighting system, the temperature or the ambient music. These spaces are usually sensorized to monitor the user's behavior and emotions and to execute the regulation interventions.

This paper presents a solution that combines the advantages of mobile applications and smart environments to regulate users' mood through the *Spotify* music. The user wears a wearable that integrates different physiological sensors. The wearable sends sensor data to a mobile application that processes them to deduce the emotions that the user is feeling during her/his daily activities. These emotions are then used to determine the user's mood. Then, when she/he arrives at home, the mobile application connects automatically with a smart system for playing music that regulates her/his mood in order to increase her/his wellbeing. These music interventions are based on the services provided by *Spotify* and considers the user's recent mood and her/his music tastes and preferences. The solution integrates portable wearables, emotion recognition systems based on artificial intelligence techniques, emotion-based music recommendation systems, content personalization tools, and an ecosystem of software services and hardware devices.

The rest of the paper is structured as follows. Section 2 presents the technological architecture of the proposed solution. Section 3 describes some implementation details of the systems integrated into the architecture. And, finally, conclusions and future work are discussed.

2 System architecture

Figure 1 shows the context and the technological elements involved in the solution. It consists of a mobile application and a set of *Spotify*-based services that regulate the user's mood through the music.

The application determines the user's emotions during her/his diary activities and the evolution of these emotions over time. The user wears a wearable device that monitors at real time different physiological parameters of her/his body, in this case an *Empatica E4* wristband [1]. These parameters are periodically sent to the application through a Bluetooth-based streaming connection. Then, the application filters the received data and deduces the emotion that the user is feeling at that moment. When the user arrives at home, the application analyses the diary emotions to determine how the user felt during the day and decides the music interventions to be applied to regulate her/his mood. These decisions are translated to voice commands that determine the type of music to be played.

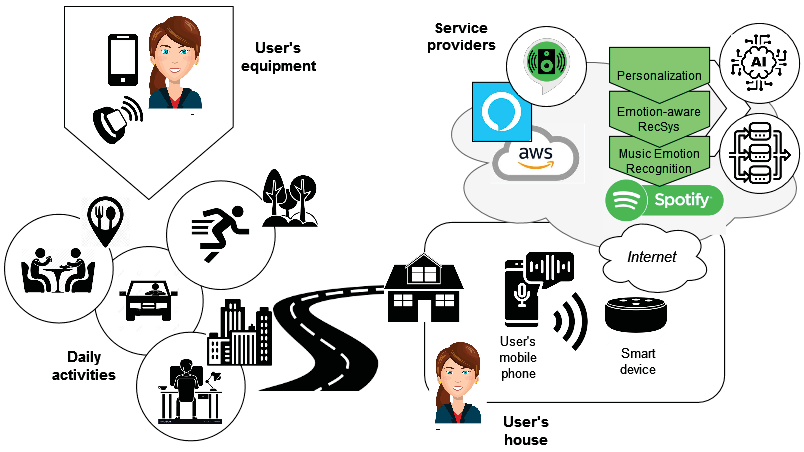


Fig. 1. Technological components involved into the solution

An *Amazon Echo* speaker is responsible for interpreting these commands and playing the most suitable songs though the *Spotify* streaming service.

The mobile application integrates three systems that provide the functionality described above: an emotion recognition system based on artificial intelligence models that deduces the user's emotions from her/his physiological data; a decision-making system based on rules that determines the mood regulation interventions, that is, the emotion to be produced in the user through the music; and, finally, a voice synthesizer that translates those interventions to commands understandable by the speaker, for example, *"Alexa, plays relaxed songs"*.

On the other hand, at the right-hand side of Figure 1 are represented the services programmed for supporting the mood regulation. A new *Alexa skill* has been programmed and installed into the speaker to provide the regulation functionality. When the skill receives a voice command, it interacts with a music recommendation system based on emotions to determine the songs to be played. The recommender selects a set of candidate songs from the *Spotify* music catalogue taking in account the emotion to be produced in the user in order to regulate her/his mood. Then, a personalization system filters and ranks these candidate songs considering the user's music preferences and tastes. Finally, the skill plays the ranked songs until a new regulation command is received.

3 Details on the system implementation

The mobile application integrates an emotion recognition system that is currently based on the analysis of an only physiological parameter, specifically, the electrodermal activity of the user's skin (EDA). Four machine learning models have been combined to determine the user's emotions from the EDA signal sent at real time from her/his wristband. We concluded as part of the research

that the application of the same machine learning algorithm to recognise the different emotions was not the best option. Therefore, a multi-model approach based on different algorithms has been applied to improve the accuracy of the recognition. These emotions are represented according to the reference model of affect proposed by Russell [7]. The different emotions felt throughout the day are then processed by an mood-recognition algorithm to create the user's mood map. This map represents the evolution of the mood over a period of time.

The user's mood map is the input of the decision making system included into the application. It integrates a rule engine that determines the type of music to play for regulating the user's mood. Different sets of rules are evaluated according the user profile. These rules have been defined though different experiments with real users (42 rules and more than 200 participants). The result of the rule-based evaluation is the emotion to be produced in the user. This emotion is then translated to a voice command that is played by the synthesizer when she/he arrives at home. A catalog of emotion-aware commands has also been defined to be interpreted by the *Amazon Echo* speaker.

The speaker's functionality requires to know the emotional response that the songs available in *Spotify* produce in the listeners. As part of the work, we have studied the relationship between these responses and the songs' audio features. An emotion recognition system based on audio features was developed to annotate *Spotify* songs from an emotional perspective. This system integrates a set of machine learning models that determine the produced emotion for each particular song and create the corresponding emotional labels. After processing all the *Spotify* catalog of songs, we obtained a database of emotionally labelled songs which is used by the music recommendation system to suggest candidate songs in response to voice commands. A detailed description of the recognition system and the recommender can be found in [6].

Finally, the personalization system customises the music experiences and, therefore, it plays a relevant role in the mood regulation process. The system automatically creates a user model that describes the user's musical preferences and tastes. This model is calculated from her/his *Spotify* listening histories and the playlists that she/he has created in the streaming provider. These data are published by *Spotify* through its service platform for developers (we are specially interested in the songs' metadata and audio features and the registered users' playlists and listening histories). Once the model has been calculated, the system applies collaborative filtering techniques to rank the candidate songs suggested by the recommender and to select the most suitable ones to regulate the user's mood.

4 Conclusions and future work

In this paper we have presented a technological architecture to promote music experiences that improve users' wellbeing. A first version of the architecture has been implemented in which the affective dimension is the cornerstone of the solution. Different emotion-aware systems have been integrated to recognise the

user's emotions, to deduce her/his mood, to determine the affective response associated to songs and to recommend songs that regulate the listener's mood. These systems are integrated with the services offered by *Spotify* providing a final solution based on its music streaming service.

As future work, we are interested in organising long-term experiments with real users to evaluated the impact of the proposal in the users' wellbeing. These experiments should be focused on certain population segments of interest. Besides, we would also like to enhance the emotion recognition models including new physiological parameters and the regulation rules and the personalization models from the feedback of the experiments.

References

1. Empatica e4 description. https://www.empatica.com/en-eu/research/e4/ (2021), accessed: 2019-11-30
2. Afzal, M., Ali, S.I., Ali, R., Hussain, M., Ali, T., Khan, W.A., Amin, M.B., Kang, B.H., Lee, S.: Personalization of wellness recommendations using contextual interpretation. Expert Systems with Applications **96**, 506–521 (2018). https://doi.org/https://doi.org/10.1016/j.eswa.2017.11.006
3. Calvo, R.A., Peters, D.: Positive computing: technology for well-being and Human Potential. The MIT Press (2014)
4. Chambel, T., Carvalho, P.: Memorable and emotional media moments: Reminding yourself of the good things! In: Chessa, M., Paljic, A., Braz, J. (eds.) Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2020, Volume 2: HUCAPP, Valletta, Malta, February 27-29, 2020. pp. 86–98. SCITEPRESS (2020). https://doi.org/10.5220/0009178300860098
5. Duvall, S., Spurlock, S., Duvall, R.: Automatic environment adjustment for emotional disabilities. In: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'17). pp. 363–364 (10 2017). https://doi.org/10.1145/3132525.3134816
6. de Quirós, J.G., Baldassarri, S., Beltrán, J.R., Guiu, A., Álvarez, P.: An automatic emotion recognition system for annotating spotify's songs. In: Panetto, H., Debruyne, C., Hepp, M., Lewis, D., Ardagna, C.A., Meersman, R. (eds.) On the Move to Meaningful Internet Systems: OTM 2019 Conferences. pp. 345–362. Springer, Cham (2019)
7. Russell, J.: A circumplex model of affect. Journal of personality and social psychology **39**(6), 1161–1178 (1980)
8. Timoney, J., O'Leary, S., Czesak, D., Lazzarini, V., Conway, E., Ward, T.E., Villing, R.: The beathealth project: Application to a ubiquitous computing and music framework. Journal of Cases on Information Technology **17**(4), 29–52 (2015)
9. Tolks, D., Sailer, M., Dadaczynski, K., Lampert, C., Huberty, J., Paulus, P., Horstmann, D.: Onya: The wellbeing game: How to use gamification to promote wellbeing **10**(2) (2019). https://doi.org/10.3390/info10020058
10. Yu, B., Hu, J., Funk, M., Feijs, L.: Delight: biofeedback through ambient light for stress intervention and relaxation assistance. Personal and Ubiquitous Computing **22**, 787–805 (08 2018). https://doi.org/10.1007/s00779-018-1141-6

Towards a resilient e-health system for monitoring and early detection of severity in hospitalized patients during a pandemic

Rodrigo S. Cañibano[1][0000−0001−6992−5421], Santino Castagno[1], Mariano Conchillo[1][0000−0002−4241−3511], Guillermo Chiarotto[1], Claudia Rozas[1][0000−0002−9876−4724], Claudio Zanellato[1], Cristina Orlandi[2], and Javier Balladini[1⋆][0000−0002−9769−7830]

[1] Universidad Nacional del Comahue, Neuquén, Argentina,
{rcanibano, stefano.castagno, mariano.conchillo, claudia.rozas, claudio.zanellato, javier.balladini}@fi.uncoma.edu.ar, guillermo.chiarotto@est.fi.uncoma.edu.ar
[2] Hospital Área Programa General Roca, Río Negro, Argentina,
cristina.orlandi@gmail.com

Abstract. During a pandemic, the traditional methodology of doctors and nurses must be optimized to allow the care of a greater number of patients without reducing the quality of care. The triage method select and classify patients in different levels of severity to achieve a correct allocation of care. Although triage can be performed through manual calculations, the process would be error-prone and considerably increase the workload on health personnel, limiting the potential benefit of the method. Our research group designed and developed *COVINDEX*, a computer system aimed at monitoring and early detection of severity in hospitalized *COVID-19* patients. It implements an existing early warning system for patients hospitalized in general wards. Our system has a complex distributed architecture that is easy to deploy and maintain, and it provides high resilience to network and hardware failures. This work discusses some details of the system architecture.

Keywords: Early severity detection · e-health monitoring · fault tolerant distributed architecture.

1 Introduction

The *SARS-COV-2* pandemic has shown the health system's lack of preparation to face this type of situation. Health centers were overwhelmed by the sudden increase in demand. Specialized human resources became scarce, aggravated by contagions from the staff themselves, which led to them temporarily ceasing to provide services.

In this context, the traditional methodology of doctors and nurses must be optimized to allow the care of a greater number of patients without reducing the

⋆ Corresponding author.

quality of care. To organize and optimize resources in this critical situation it is possible to use the triage method. This method consists of the selection and classification of patients in different levels of severity to achieve a correct allocation of care. Triage was applied early in China [3] during the health emergency of *COVID-19* patients through the adaptation of one of the best known and validated *Early Warning Systems (SAT)* [2]. This allows staff performance to be improved, reducing controls in stable patients, increasing them in seriously ill patients, and reducing unexpected hospital mortality in general hospital areas.

Although triage can be performed through manual calculations, the process would be error-prone and considerably increase the workload on health personnel, limiting the potential benefit of the method.

To improve the quality of care [4], our research group designed and developed *COVINDEX*, a computer system aimed at monitoring and early detection of severity in hospitalized *COVID-19* patients. It implements the early warning system published in [1] for patients hospitalized in general wards, based on the knowledge of expert health personnel. It also allows the collection of data for later improvement of the prediction using machine learning techniques.

COVINDEX is a clinical decision support system that can be interconnected with a digital medical record. Our system has a complex distributed architecture that is resilient (the system continues to provide services in the event of network and device failures), and simple to deploy and maintain (it does not require the use of standard servers in the hospitals). To the best of our knowledge, no other systems exist with this properties.

2 Current development

COVINDEX performs an automatic real-time analysis of the clinical data of patients: vital signs, comorbidities, laboratories, and X-ray reports of the patient. Patients are classified according to four severity levels (according to what is expected to develop after 24 or 48 hours): low, moderate, high and critical (needs intensive care). The system issues alerts to health personnel if there is a change in severity, and organizes and coordinates the work of doctors and nurses in a patient-centered care process.

2.1 Architecture overview

The architecture design must meet the following characteristics:

- Take into account that hospital wards in Argentina are highly diverse in terms of the availability and reliability of their Internet connection. Health personnel must be able to record data related to patients in the ward where they are assigned, and the system must provide early warnings, even in the event of an interruption in the connectivity of local devices with system components located outside the hospital. Data consistency must be maintained, even after a communication interruption between system components.

- It should not require a standard server in the hospitals to simplify the deployment and maintenance of the system.
- A global and historical repository of data is required for the generation of statistical information and the extraction of new knowledge.

The system architecture can be summarized in three computing levels:

Edge: mobile devices used by doctors and nurses. They allow data to be entered into the system and the results to be observed. It can be directly connected to the fog level through a local area network or cloud level through the Internet. Cloud level connection is preferred to reduce overhead and power consumption of fog level devices.

Fog: mobile devices located inside the hospital. They maintain the primary copy (master copy) of a part of the global database. If connection to the cloud level is lost, it allows edge devices to be served through direct connections established over the local area network. This device can also be simultaneously used as an edge level device.

Cloud: maintains a complete and historical copy of all fog level device databases, and is a primary copy of hospital administrative data, and data shared between all hospitals. Also, it can reduce the overhead of fog level devices by taking care of edge devices.

2.2 Detailed description of the architecture

Three different types of domains are defined (shown as circles in figure 1):

Users domain: which contains the data of all users that are relevant to all hospitals, such as username, email, and the password data. There is only one instance of this domain.

Hospital domains: each one contains the specific data of a particular hospital, such as the list of users who are employees of the hospital, wards and beds. There is one instance of this domain per hospital.

Island domains: each one contains data specific to a group of wards of a particular hospital, such as vital signs, laboratories, and X-ray reports of patients. There are multiple instances of this domain on each hospital. Which wards belong to an island are defined by the connectivity between them and the capabilities of the fog device that manages them.

Each domain has a single leader device that holds the primary copy of the data, and client devices that hold replicas. When a device wants to store a data record for a specific domain, it must send a request to the leader of that domain. When the leader stores the record on its database, it sends a copy to all client devices in that domain, including the device that made the request. The leader might process the inserted record in conjunction with the historical data of the patient to generate new records (for example, related to triage).

To keep the devices up to date even after the connection to some other system component is lost, each domain instance has a virtual clock (represented as an

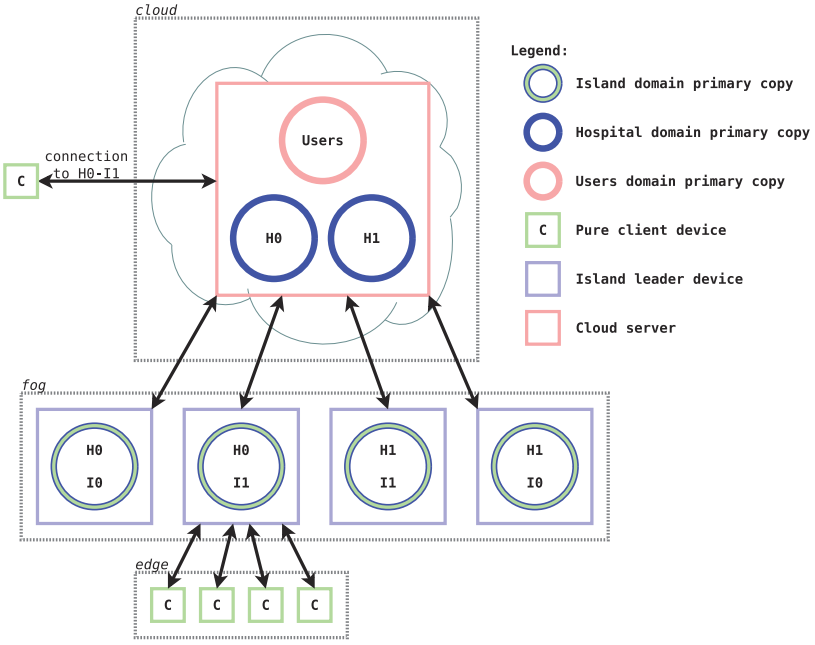


Fig. 1. Detailed architecture of the system

integer) that counts the number of transactions (insertions of new records). We call this virtual clock *sync_id*. Each record has the *sync_id* of the domain as an additional key. To keep a complete history of the inserted data, the records are never modified or deleted, and the most current version of the record is distinguished by the value of the logical clock. When a client device reconnects, it can ask the domain leader device for records with *sync_id* greater than the last one the device has stored.

Leader devices can be clients (maintain replicas) of each other, which allows them to serve as caches for other domains. Furthermore, they can be used by clients as bridges to leaders from other domains. This allows a leader device to reduce the workload on leaders with more limited capabilities and, if the network becomes disconnected, it can provide clients (still connected to the leader) with the latest data records stored before failure (properly indicating that this data may be out of date).

The current system defines two types of leaders and a pure client type (shown in figure 1 as squares):

Cloud server: is the leader of the users domain and of all the hospital domains, it is also a client of all the island domains. It maintains a complete and historical replica of all domains.

Island leader devices: these are fog level devices. Each one is a leader of an island domain, a client of a hospital domain, and a client of the users domain.

They are physically located within the hospital. Upon interruption of connectivity to the cloud server, clients can continue to operate by interfacing with the island leader device via a local area network.
In addition, each time a new record of a patient is inserted, a triage is performed (also using historical data of the patient), which determines a level of severity. If the severity of the patient has changed, an alert is issued.

Pure client devices: these are the edge level devices. They are clients of a single island domain, a single hospital domain, and the users domain. They keep a replica of these domains. In the event of an island leader device failure, one of the pure client devices can be promoted to island leader device.

3 Next steps

Since the database is distributed within the hospital and the devices can be easily replaced, the current system is highly resilient to connectivity and device failures (island leaders and pure clients). Meanwhile, the cloud server centralizes its services. To increase fault tolerance of this component, a spare node should be installed and a mechanism that allows the service to be automatically restored in the event of a failure should be implemented.

The *Ministry of Health of the Province of Río Negro, Argentina* is interested in deploying the system in its public hospitals. The system will be expanded for the following purposes: (1) Timely referral of critical patients from hospitals without intensive care units; (2) Unify the care process for hospitalized patients, and increase the quality of care; (3) Obtain statistical information from all hospitals in the province for decision making; (4) Extend the use to Non-COVID-19 patients.

Data is currently recorded into the system manually, and therefore data flow is limited. However, it may be desirable in the future to connect sensors at the edge level. For this the architecture will need to be tailored to support this new feature.

We still need to evaluate the system performance under a realistic load and improve the security of the data stored on edge and fog devices.

References

1. Huespe, I., Carboni Bisso, I., Gemelli, N.A., Terrasa, S.A., Di Stefano, S., Burgos, V., Sinner, J., Oubina, M., Bezzati, M., Delgado, P., et al.: Design and development of an early warning score for covid-19 hospitalized patients. MEDICINA (Buenos Aires) **81**(4) (2021)
2. Lee, J.R., Kim, E.M., Kim, S.A., Oh, E.G.: A systematic review of early warning systems' effects on nurses' clinical performance and adverse events among deteriorating ward patients. Journal of patient safety **16**(3), e104–e113 (2020)
3. Liao, X., Wang, B., Kang, Y.: Novel coronavirus infection during the 2019–2020 epidemic: preparing intensive care units—the experience in sichuan province, china. Intensive care medicine **46**(2), 357–360 (2020)
4. Rocco, C., Garrido, A.: Seguridad del paciente y cultura de seguridad. Revista Médica Clínica Las Condes **28**(5), 785–795 (2017)

Recommendations for Resilient App Development

Leonardo Corbalán[1[0000-0001-9026-8059]], Juan Fernández Sosa[1[0000-0002-0482-3392]], Fernando Tesone[1[0000-0001-9499-3127]], Sebastián Dapoto[1[0000-0001-7593-0198]], Federico Cristina[1[0000-0003-3838-417X]], Pablo Thomas[1[0000-0001-9861-987X]] and Patricia Pesado[1[0000-0003-0000-3482]]

[1] Computer Science Research Institute LIDI (III-LIDI)
School of Computer Science, National University of La Plata,
La Plata, Buenos Aires, Argentina
{corbalan, jfernandez, ftesone, sdapoto, fcristina, pthomas, ppesado}@lidi.info.unlp.edu.ar

Abstract. Applications for mobile devices tend to face a greater number of adverse situations derived from the very nature of mobility. In this context, improving resilience becomes relevant. Here we present some recommendations, based on preliminary studies, to increase resilience to the performance drop and power consumption increase caused by increased computational load.

Keywords: Resilient Mobile Applications, Performance, Battery Consumption, 3D Mobile Applications, Mobile Applications with High Computational Load.

1 Introduction

Etymologically, the term resilience comes from the Latin *resilīre*, which means 'to jump back, to bounce back', 'to withdraw' [1]. Applied to different areas of knowledge and human activities, this term has been redefined in different ways, incorporating nuances specific to each discipline. In biology, it refers to the capacity of a living being to adapt to a disturbing agent or adverse situation. In physics, it describes the capacity of a material, mechanism or system to recover its initial state when the perturbation to which it had been subjected has ceased. In psychology, resilience is the capacity of a person to adapt positively to adverse situations, overcoming the trauma they may have caused.

In engineering disciplines, resilience is understood as the capacity of a system to resist and recover from a failure in some critical component in order to maintain its operation and services in an acceptable manner [2]. The amount of recovered functionality and the time invested in the recovery are relevant parameters to determine the degree of resilience of a system in a given adverse situation. These statements also apply to software systems.

Software failures can arise from unforeseen situations, adverse circumstances, random events or malicious attacks. In the latter case, the ability of a system to withstand a cyber-attack is known as cyber resilience. Regardless of the origin of the failures, resilient software will be able to recover to a degree and period of time previously determined according to the requirements and particularities of the system.

In the case of software applications for mobile devices (apps), the resilience analysis must consider the very nature of mobility. This confers certain distinctive features that expose apps to a set of potentially adverse situations that are not present in other types of applications.

In this short paper we present a set of preliminary results from which we derive a series of recommendations for the development of apps that are resilient to the drop in performance and the increase in energy consumption caused by the increase in computational load.

The rest of this paper is organized as follows: section 2 describes adverse scenarios and strategies to improve app resilience, section 3 presents some results on app performance and energy consumption, section 4 presents a series of recommendations to increase app resilience, and finally, section 5 presents conclusions and future work.

2 Resilience in Apps

The development of applications for mobile devices must consider how to increase the degree of resilience of apps in order to cope with a greater number of stress situations to which they are often subjected. Battery drain, total or partial loss of connectivity that comes with mobility, increased risk of cyber attacks, partly due to the typical behavior of mobile device users, are just a few examples of such stress situations.

It is possible to recommend the implementation of a series of strategies to increase the degree of resilience depending on the type of adversity to which an app may be subjected.

2.1 Resilience of Apps to Connection Loss

It is common for apps to have a strong dependency on Internet connection to access and communicate with services deployed in the cloud. To build resilient apps, connection failures should not be considered exceptional and the offline experience should be considered as just another state of the application. In this way, when a connection is lost, the app will continue to operate with some limitation, but avoiding crashes, blank screens or alert messages that generate frustration and abandonment by users [3].

Upon loss of connection, the app must be able to capture the user's intent, executing the required action as far as possible and completing it later when the connection has been re-established. This type of design is referred to as "offline-first" solutions [4].

While there is no connection to the network, in the absence of the most recent data, the local memory of the device will provide the data obtained in the last stable connection. PWA, Progressive Web Applications, are mobile web applications that use this methodology to generate an offline experience. For this purpose, web technologies such as Service Workers and caching, among others, are used [5].

2.2 Resilience of Apps in the Face of Connection Instability

In case the connection is unstable, resilient behavior can be achieved by implementing two well-known design patterns: the Retry Pattern [6] and the Throttling Pattern [7].

Retry Pattern allows handling transient failures in the connection to a network service by retrying a failed operation. This strategy is based on the assumption of two premises: i) failures are of short duration and ii) the previously failed request may succeed in a subsequent attempt. It is necessary to define the maximum number of attempts to be made and the waiting time between each of them, being able to opt for a fixed or incremental waiting time.

On the other hand, Throttling Pattern consists of relaxing some requirements to reduce the demand for resources while avoiding compromising response times or the overall performance of the app. This allows the system to continue to function, in a suboptimal manner, in the face of difficulty in fully accessing a resource. Some actions to take are: limit the number of requests per user, keep essential services in the first instance and prioritize the distribution of requests.

2.3 Resilience of Apps against Cyber Attacks

Resilience to cyber attacks is known as cyber resilience. It is defined as the ability to anticipate, resist, recover from, and adapt to attacks or compromises in systems that use cyber resources [8]. Compromised systems must absorb the impacts of attacks, and respond quickly and flexibly to ensure the operational continuity of their critical components.

In recent years, mobile devices have become an essential tool for the operation and activities of companies, but, at the same time, they represent a new medium through which criminals can attack and access a company's resources. Malware attacks, social engineering, access to insecure wireless networks, weak passwords, incorrect use or careless user behavior are just some of the potential risks associated with the use of mobile devices in companies. Loss or theft of the device is also a security issue that can facilitate access to company resources by unauthorized third parties.

Some recommendations for the development of cyber resilient mobile apps [9] are: i) web services communication should use SSL/TLS (Secure Sockets Layer / Transport Layer Security), ii) URLs launched by an application should use the HTTPS protocol, iii) encrypt the information handled by the application, iv) the number of login attempts should be limited and v) development frameworks should be updated.

2.4 Resilience of Apps to Increased Computational Load

A high computational load increases energy demand and affects battery life. In addition, depending on the computational capacity of the device, it can delay app response time, which is perceived as a performance drop that degrades the user experience. As an example, consider 3D mobile applications whose processing demands can easily exceed the capacity of the devices on which they run.

In order to increase resilience to performance loss and rapid battery depletion, it is necessary to monitor response speed and power consumption in real time. In this way, measures can be taken to reduce the computational requirement (Throttling Pattern), extending the device's battery life and improving the app's response time.

3 Study of Performance and Power Consumption in Apps with High Computational Load

For this analysis it is necessary to consider the development frameworks used because they impact the efficiency with which the apps consume or use the available resources. The performance and power consumption of apps built with the following development frameworks were studied: Android SDK (native), Cordova, Titanium, NativeScript, Xamarin, and Corona. [10]. The best response times and most efficient power consumption, during processing-intensive tests, were obtained by applications developed with Cordova, Titanium, and NativeScript (see Table 1).

Table 1. Intensive processing app

Framework	Power (mWh)		CPU charge (%)		Duration (s)	
	$\overline{E}$	S_E	$\overline{C}$	S_C	$\overline{T}$	S_T
Cordova	1.597	0.136	35.924	2.571	8.467	0.679
Titanium	1.692	0.096	37.480	2.395	8.355	0.643
NativeScript	1.792	0.176	33.357	2.217	9.109	1.789
Xamarin	3.036	0.185	32.072	1.768	17.891	0.973
Android SDK	3.463	0.149	32.468	1.332	18.568	2.938
Corona	7.304	0.189	44.347	54.793	38.877	1.492

With respect to 3D applications for mobile devices, 2 of the most currently used development frameworks were evaluated: Unity and Unreal Engine [11]. The characteristics of a 3D application that have the greatest impact on performance and energy consumption on a mobile device were identified. It was observed that applications developed with Unity consume less energy, but at the same time present a greater performance degradation, compared to Unreal Engine, when the complexity of the scene to be visualized increases drastically (see Table 2).

Clearly, the factors that most affect performance and power consumption in both frameworks are the display of complex objects (high number of polygons) and the simultaneous use of different particle systems (e.g. smoke, sparks, explosion, etc.).

Table 2. Performance and consumption of 3D mobile applications

Scene characteristics	Performance as scene complexity increases of the scene with increasing number of objects (FPS)		Consumption (mAh)	
	Unity	Unreal	Unity	Unreal
Simple objects	60 60 60 60 60 47 16 04	60 61 60 58 40 28 10 08	19.8	51.53
Complex objects	40 34 17 08 05 05 04 04	08 08 08 08 08 08 08 08	27.18	61.06
Lights and shadows	60 60 60 60 59 40 15 04	60 60 60 59 42 30 10 08	20.73	37.4
Textures	60 60 60 60 60 47 13 04	60 60 61 56 37 29 12 08	20.23	47.48
Particles	60 60 46 14 04 03 00 00	19 19 16 11 08 08 08 08	15.94	21.78
Physics	60 60 60 60 60 45 03 03	60 60 60 58 42 30 10 08	18.36	18.65

4 Recommendations. Preliminary Results

A series of recommendations are presented here to improve the resilience of apps to performance drops (longer response times) and increased energy consumption caused by increased computational load.

The first aspect that must be taken into account to improve resilience is the app's ability to offer resistance to the adversity it faces. For this purpose, development frameworks should be used to obtain apps that better manage energy consumption and computational load.

It is recommended to use the multiplatform development frameworks Cordova, Titanium and NativeScript for general-purpose applications with high computational load.

For the development of 3D applications with medium or low complexity scenes, it is recommended to use Unity. Unreal Engine can process more powerful and realistic graphics than Unity but with higher power consumption. For the case of very complex scenes, Unreal Engine should be considered because the performance of apps generated with this framework does not degrade as much as the performance of apps developed with Unity.

The second important aspect to improve resilience is to increase the capacity to recover from a stressful situation. In this sense, an effective strategy is to use the Throttling Pattern. For this it is necessary to monitor the response speed and energy consumption in real time. In 3D applications this can be achieved by reducing the complexity of the scene to be visualized. According to the data analyzed, and due to the impact they have on performance and energy consumption, it is recommended to reduce some or several of these factors that affect the complexity of the scene: i) Number of polygons in the scene, ii) particle systems to visualize, iii) dynamic lights and shadows of the objects and iv) complex textures or materials of the objects.

5 Conclusions and Future Work

In this short paper we have analyzed, from a resilience perspective, some experiments on the impact of development frameworks on the energy efficiency of the apps built.

From this analysis, a series of recommendations have been made to improve the resilience of apps to the performance drop and energy consumption increase caused by the increased computational load. Recommendations were made for the development of general-purpose applications and for 3D mobile applications.

As future work, we plan to extend the experiments to other frameworks, and also to study the behavior of the developed apps in the event of partial or total loss of connectivity. We also intend to incorporate the analysis of using local DBs with cloud synchronization to mitigate the loss of connectivity.

Finally, we will seek to expand the analysis of resilience aspects in 3D Apps, from the incorporation of Virtual Reality (VR).

References

1. Real Academia Española. Resiliencia. En Diccionario de la lengua española, 23.ª ed., versión en línea. Recuperado de https://dle.rae.es/resiliencia.
2. Murray, A., Mejias, M., & Keiller, P. (2017). Resilience methods within the software development cycle. In Proceedings of the International Conference on Software Engineering Research and Practice (SERP) (pp. 62-65). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
3. Lazar, J., Jones, A., Hackley, M., & Shneiderman, B. (2006). Severity and impact of computer user frustration: A comparison of student and workplace users. Interacting with Computers, 18(2), 187-207.
4. Vanhala, J. (2017). Implementing an Offline First Web Application.
5. Aguirre, V., Delía, L., Thomas, P., Corbalán, L., Cáseres, G., & Sosa, J. F. (2019, October). PWA and TWA: Recent Development Trends. In Argentine Congress of Computer Science (pp. 205-214). Springer, Cham.
6. Retry pattern - Azure Architecture Center | Microsoft Docs https://docs.microsoft.com/en-us/azure/architecture/patterns/retry.
7. Throttling pattern - Azure Architecture Center | Microsoft Docs https://docs.microsoft.com/en-us/azure/architecture/patterns/throttling
8. Ross, R., Pillitteri, V., Graubart, R., Bodeau, D., & McQuaid, R. (2021). Developing cyber resilient systems: a systems security engineering approach (No. NIST Special Publication (SP) 800-160 Vol. 2). National Institute of Standards and Technology. available free of charge from: https://doi.org/10.6028/NIST.SP.800-160v2r1
9. Jaramillo, H. D., Romero, G. K., & Ramos, C. (2021). Framework de seguridad para desarrollo de aplicaciones móviles y su aporte a la CiberResiliencia. Revista Ibérica de Sistemas e Tecnologias de Informação, (E42), 452-468.
10. Corbalán, L., Thomas, P., Delía, L., Cáseres, G., Sosa, J. F., Tesone, F., & Pesado, P. (2019, June). A study of non-functional requirements in apps for mobile devices. In Conference on Cloud Computing and Big Data (pp. 125-136). Springer, Cham.
11. Cristina, F., Dapoto, S. H., Thomas, P. J., Pesado, P. M., Perez Altamirano, J., & De la Canal Erbetta, M. (2020). Aplicaciones Móviles 3D: un estudio comparativo de performance y consumo de energía. In XXVI Congreso Argentino de Ciencias de la Computación (CACIC)(Modalidad virtual, 5 al 9 de octubre de 2020).

Computational Modeling and Simulation

ABM simulation focused on urban mobility

Antonio González Cuevas, Remo Suppi

Universitat Autònoma de Barcelona, Dept. de Arquitectura de Computadores y Sistemas Operativos, 08193, Cerdanyola (Barcelona), España
{toni.gonzalezcuevas@gmail.com, remo.suppi@uab.cat}

Abstract. The objectives of this research is to solve, through high-performance simulation, different mobility problems in cities with mixed traffic (public and private transport, emergencies, bicycles, scooters). To achieve these objectives, the research is oriented to develop new ABM models (Agent Based Models), integrated with Geographic Information Systems (GIS) in order to provide answers to different situations and mobility scenarios in big cities. We want to analyze different episodes and strategies (real or possible) to solve questions based on 'What if?' with variables such as mixed traffic, bicycles exclusive lanes, reduction of private vehicles lanes, among others. This work shows the first phase of ongoing research focused on the city of Barcelona, but the proposed methodology and models can be quickly extended to other cities. For this, we work with data captured in real traffic analysis processed with vehicle recognition algorithms based on AI (artificial intelligence) and the results are used as the basis for the ABM model based on GIS. The simulation is carried out by integrating initial real data with those generated by the simulation in order to analyze the mobility and their interactions.

Keywords: Agent Based Modeling, urban mobility simulation, real-time data, high performance computing.

1 Introduction

Simulation models aim to analyze different aspects of reality, capturing capabilities and behavior. These models allow knowing possible real scenarios but before they occur. Methods to analyze real scenarios are common in different disciplines for training and decision making, but these methods can be very complex or very costly (are used, for example, in evacuation of buildings, terrorist alerts, floods, etc.).

One of the important problems of the simulation is that, habitually, the models are proposed for a set of initial conditions but these models are disconnected from reality when the simulation advances. To solve this problem, large amounts of simulations must be carried out to analyze the different possible scenarios. In urban environments and traffic simulations, other alternatives are possible, such as having real data to 'reorient' the simulation to generate different analyzes and extract knowledge. [1, 3, 4]

For example, a use case could be simulating one of the streets with the highest traffic density at Barcelona, using real data and 'causing', through simulation, a multiple accident to analyze the situation on this street and its surroundings. With this information, it would be possible to determine the origin and best route for the

emergency teams based on the minimum arrival time at the accident place in a collapsed environment with a very high density of vehicles.

For this, it is necessary to have reliable real data that can be injected into the simulator and to develop methodology and models that, focused on a context of mobility in cities, allow us to draw conclusions and knowledge using simulations. This type of model will allow the generation of decision support tools based on data, not only for critical situations, but also in the forecasting of complex urban mobility models such as, for example, 'what is the impact of adding a bicycle lane on a high density street? If the maximum speed is reduced by 40%, how does it affect mobility? or 'the mixed traffic (cars, public transport, motorcycles, bicycles, skateboards) is possible on high density way?. [3,6]

This research proposes to provide answers to these situations using high-performance simulation based on agents, considering real data to analyze mobility flows and interactions between different types of vehicles, as well as the conditions and consequences of these interactions. The strategy to generate open simulations (no only for a specific city only) is to use GIS systems to adapt the environment to the simulation framework. With this, the ABM simulation + GIS environment + real data will allow us to respond to different real situations and scenarios in a city or complex mobility environments to analyze different episodes and strategies (real or fictitious).

2 Objectives and methodology

The main objective of this project is based on the development and validation of urban mobility maps using agent-based simulation models. The requirements are: infrastructure by GIS-based models, mixed and priority traffic, effective validation on real data, Decision-Making-Support (DMS)-oriented, ABM-based design that can be migrated to Cloud and HPC environments when higher performance is required.

For this, the following design aspects have been considered: GIS model centered on Barcelona, incremental design (phases), multi-lane support, regulated traffic (traffic lights), vehicles of different types with adjustable parameters (initially: category, maximum speed, density and priority). In second term, the incorporation of priority vehicles, integration with Open-Data (traffic, congestion points, greatest accident rate points, etc.), scalability at the city level /region, migration to HPC for large city models will be considered.

An incremental phased development has been used as a methodology. This methodology will allow consolidating the objectives centered on 4 major tasks: 1) collection and extraction of real data, 2) integration of GIS data in agent-based simulation environments, 3) simple mobility models based on real data on GIS infrastructure, 4) multilane models with mixed traffic centered on an urban road with a high density of controlled traffic and its surroundings.

3 Current & future developments

Task 1: Collection and extraction of real data.

For this task, a framework based on artificial intelligence was developed for the recognition of vehicles from a real video of the road of interest and its surroundings. This environment allows us to identify the different vehicles on the road, obtaining the number of vehicles of each lane, direction, speed and distance. The initial work has been developed based on the DeepSORT and YOLOv5 [2] libraries with a first analysis focused on cars on a high density Barcelona street (Gran Vía de les Corts Catalanes) to obtain data on speed distributions and distances between cars. The following figure shows the developed data collection and recognition environment.

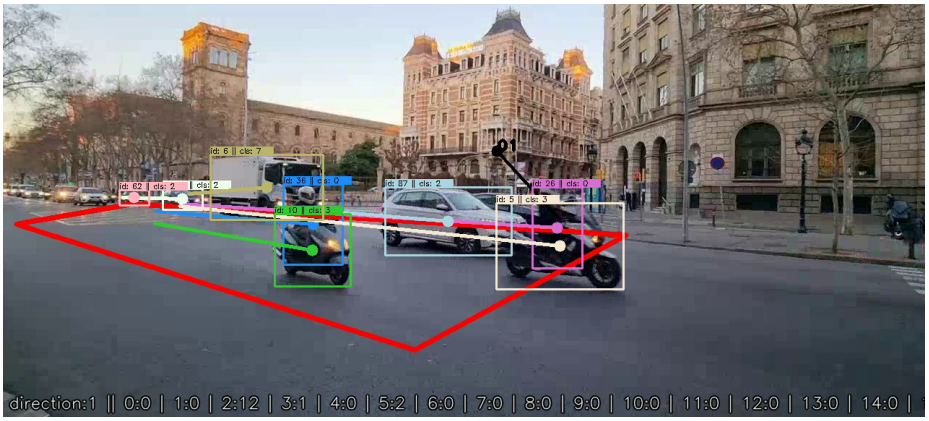


Fig. 1. Identification of vehicles in Gran Vía de les Corts Catalanes street (Barcelona).

These data will be used to define the preliminary ABM simulation model.

Task 2: integration of GIS data in agent-based simulation environment.

The development of this task focuses on integrating a GIS map on an agent-based simulation environment in order to draw the GIS streets within the simulation environment. As ABM simulation environment, NetLogo [7] has been selected since it supports GIS extensions [8] and the models can be migrated to an HPC environment without problems. Figure 2 shows the integration of the map of Barcelona in the simulation environment where the roads (blue lines) are active part of the simulation framework (gray lines are visualized as background image).

Task 3: simple mobility models based on real data on GIS infrastructure.

With the data and GIS integration developed in the previous tasks, a first preliminary simulation model based on ABM has been developed with the following characteristics: multi-agent environment (cars) based on NetLogo, uniform speed distribution based on real data, constant vehicle density, single lane (with contention), speed management. The figure 3 shows the visualization of the Netlogo environment with the cars and the different parameters. As preliminary results, average minimum

speeds is near to 20 km/h and average maximum speeds is 52 km/h are obtained similar to real data used.

Fig. 2. GIS model integrated in Netlogo

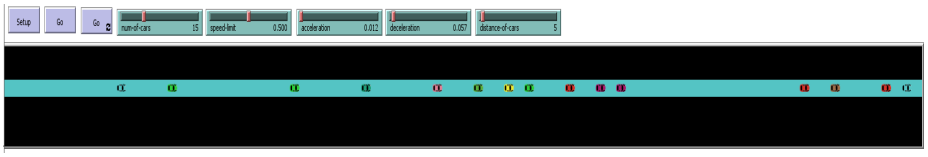


Fig. 3. visualization of single line simulation in NetLogo

In this first simple simulation, has been possible to verify that in order to obtain fluid traffic, the intensity of the traffic is the main factor and how the acceleration/deceleration factors have a direct impact on the levels of retention generated. As a next step in the investigation is the complete simulation model to include the different requirements proposed in the objectives.

4 Conclusions.

This line of research is an ambitious project that aims to integrate real data with agent-based simulation and GIS maps. Considering the complexity of some aspects of the problem, a development by phases has been chosen. The current results of the project are visualized as: capture and identification of the parameters of the streets to be analyzed, the integration of a GIS environment with an agent-based simulation environment and a preliminary simple agent-based simulation model. From the simulated model, the first preliminary results have been obtained that validate the real data, where factors such as traffic intensity, speed and as well as vehicle acceleration and deceleration are critical elements in city traffic.

Acknowledgments: This research has been supported by the Agencia Estatal de Investigacion (AEI), Spain and the Fondo Europeo de Desarrollo Regional (FEDER) UE, under contract PID2020-112496GB-I00 and partially funded by the Fundacion Escuelas Universitarias Gimbernat (EUG).

References

1 Ki Yong Lee, YoonJae Shin, YeonJeong Choe, SeonJeong Kim, Young-Kyoon Suh, Jeong Hwan Sa, and Kum Won Cho. Design and implementation of a data-driven simulation service system. EDB'16, 77–80, USA, 2016.

2 YoloV5 https://github.com/ultralytics/yolov5 https://github.com/ZQPei/deep_sort_pytorch

3 Vallati, M.; Magazzeni, D.; De Schutter, B.; Chrpa, L.; McCluskey, T.L. Efficient Macroscopic Urban Traffic Models for Reducing Congestion: A PDDL + Planning Approach. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), pp. 3188–3194, USA. 2016.

4 Maggi, E., Vallino, E.. Understanding urban mobility and the impact of public policies: The role of the agent-based models. Research in Transportation Economics. 55, pp 50-59. 2016.

5 A. Gaudiani, E. Luque, P. García, M. Re, M. Naiouf, and A. De Giusti. Computing, a powerful tool for improving the parameters simulation quality in flood prediction. Procedia Computer Science, 29, pp 299-309, International Conference on Computational Science. 2014.

6 Hu X. Dynamic data-driven simulation: Connecting real-time data with simulation. In Yilmaz L., editor, Concepts and Methodologies for Modeling and Simulation. Simulation Foundations, Methods and Applications. Springer, 2015

7 Wilensky, U. Netlogo. User Manual. https://ccl.northwestern.edu/netlogo/docs/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. 2022.

8 NetLogo Gis Extension. https://ccl.northwestern.edu/netlogo/docs/gis.html. 2022.

Evaluation of the Hospital Emergency Service, in critical situations, through the inclusion of Resilience and the Sustainable Development Goals

Mariela Rodriguez[1], Francesc Boixader[2], Alvaro Wong[3], Dolores Rexachs[4] and Emilio Luque[5]

[1] Facultad de Ingeniería, Universidad Nacional de Jujuy, Argentina
[2] Escoles Universitaries Gimbernat (EUG), Computer Science School, Universitat Autonoma de Barcelona, Sant Cugat del Vallès, Barcelona, 08174, Spain
[3, 4, 5] Computer Architecture and Operating Systems Department, Universitat Autónoma de Barcelona Campus UAB, Edificio, Bellaterra, Barcelona 08193, Spain
[3]mariela.rodriguez@fi.unju.edu.ar
[2]francesc.boixader@eug.es
[3]alvaro.wong@uab.es
[4]dolores.rexachs@uab.es
[5]emilio.luque@uab.es

Abstract. Simulation, in health services, has become an important tool that has made it possible to replicate real scenarios that have a critical degree and thus have reliable information to obtain knowledge about the variables managed in an emergency. In the activity of a Hospital Emergency Service (HED), we can find active agents such as doctors, nurses, patients and passive ones such as medical rooms, beds, among others, interacting between them. HEDs have been overwhelmed by catastrophic events, such as those induced by COVID 19, in the last two years. In these situations, the availability of resources, due to the considerable reduction in health personnel and the excessive arrival of patients for care, implies that the emergency service has become inoperative. The recovery of this essential service is fundamental and a priority. For this reason, it is necessary to analyze the period of time for departments, after a catastrophe, to reach a stable state that allows them to adapt to the new workload, that is, to analyze how resilient the ED is. The present work shows the proposal of a research project for the modeling and simulation of an ED that supports the design and evaluation of a resilient service considering the infrastructures, resources and services that occur in an ED. The model must consider the organization and infrastructure capacity to support and recover from extreme situations, in order to provide and maintain its operation and services in an adequate manner. To achieve this objective, the normal service index and the ability to resist, restore and evolve in the face of future risks and threats will be measured. The design of the model will be integrated and participatory, and it will focus on people and communities, governed by the principles of sustainable development goals regarding Health and Well-being, Gender Equality, Industry, Innovation and Infrastructure.

Keywords: Simulation, hospital emergency services, resilience.

1 Introduction

The continued enlargement in the interconnection at the economic, social, and technological levels has some repercussions on the measurements performed on health and well-being. The migration of people, natural disasters, and pandemics require that the measures taken on health be a priority and of global nature. A medical emergency service (ED) is the place where patients arrive to be treated for situations in which their lives may be at risk. According to this situation, the time that a patient spends in this space should be minimal in order to prevent the disease from advancing and the risk of catching another disease.The operation of the HEDs is a set of material and human resources adapted to an inflow of patients. In that way, the enormous increase in the number of patients causes the saturation of the service to the point of becoming disrupted. Considering that the emergency service is a priority service therefore it is necessary to be activated again as soon as possible.

At the IV World Symposium on Health Systems Research, the Director of the Pan American Health Organization highlighted the need to build stronger and more resilient health systems in global public health [1]. Resilient health systems are informed, evidence-informed, responsive, predictable, complex, adaptive, robust, integrated, and participatory, and centered on people and communities [2].

At the global level, frameworks and agreements have been registered to improve the well-being of people, such as the Sendai Framework for Disaster Risk reduction 2015-2030 signed at the III World Conference of the United Nations in 2015.

The Sendai Framework is an agreement that aims to work to mitigate the risks generated by disasters at the different levels of government organization, such as exposure, vulnerability and hazard characteristics. This agreement seeks to mobilize investments to strengthen the resilience of health infrastructure through international cooperation and working alliances. Thus, achieve the reduction of disaster risks and the losses caused by them in lives, means of subsistence and health of people, companies and communities [3]. The priorities for action that were established in the Framework were:

Priority 1: Understand disaster risk.

Priority 2: Strengthen disaster risk governance to manage disaster risk.

Priority 3: Invest in disaster risk reduction for resilience. Public and private investments are needed for disaster risk prevention and reduction, and to increase health resilience, in communities and countries.

Priority 4: Increase advance preparedness for disasters in order to provide an effective response and “build back better” in the areas of recovery, rehabilitation and reconstruction. It is important to empower women and people with disabilities to promote gender equity and access to services.

This framework revealed that local and national health systems continue to be fragile and susceptible to disaster events worldwide.

On the other hand, in 2015 the United Nations placed on the world agenda to strengthen the Sustainable Development Goals (SDGs), which allow working on actions to end poverty, protect the planet and guarantee that by 2030 all people enjoy peace and prosperity. [4]

SDG 3: Health and Well-being is necessary for humans to enjoy good health to meet this goal. The advancement of other SDGs is required, such as ending poverty and reducing inequalities. This SDG aims to strengthen early warning, risk reduction and its management for national and global health.

SDG 5: Gender equality, seeks to end all forms of discrimination against all women and girls, such as those that occur in the public and private spheres.

SDG 9: Industry, innovation and infrastructure requires reducing the digital divide to guarantee equal access to information and knowledge. As a goal, reliable, sustainable, resilient and quality infrastructures for human well-being must be developed.

The analysis of health services when extreme situations occur is difficult to measure in reality. Due to this, it is essential to use simulated models. Simulation, for health services, has become a tool that has made it possible to replicate real scenarios that have a critical degree and, thus, have reliable information to obtain knowledge about the variables handled in an emergency system. The challenge of this project is that the simulation model generated, is measured with real data and is calibrated for a determined purpose. The model must include the priorities of the Sendai Framework and the SDGs of health and well-being, gender equality in care and in the workforce, so in that way, it can design a resistant and reliable infrastructure.

The research project will be carried out with the help of a simulator used in previous works of the High Performance Computing Research Group for Efficient Applications and Simulation. The development line is called Simulation and optimization of Emergency Services in Hospitals (Smarter Health Services) where the related research is:

· Simulation of the Emergency Department of a Hospital: a computational approach improved through Individual Oriented Modeling (MoI) techniques. Where a useful simulation model for making operational decisions has been designed and developed. Through the simulator, information is available to be used for decision-making by those responsible for the emergency services [5]..

· Programming the admission of nonthreatening patients in a Hospital Emergency Service. A method proposed to try to reduce the total time of stay of patients in the service, through a model for planning the arrival of non-critical patients to it. The basis of the model is on the detailed characterization of the system in terms of its care capacity and the number of patients in care each hour in a dynamic way. [6].

2 Objetives

The recovery of an essential service is fundamental and a priority. The design of a Hospital Emergency Service (US) model is suggested, considering the infrastructure and management of resilient resources, which guarantees its durability and renewal in critical situations.

As specific objectives:

· Predict the behavior of the variables under different conditions and critical situations, carrying out inference processes on them and thus analyze the behavior, for example in the face of extreme resource reductions due to strikes, pandemics or natural disasters.

· Increase resilience in health infrastructure according to priority 3 of the Sendai Framework

· Consider in the HED management model the SDGs of Health and Well-being, Gender Equality and Industry, Innovation and Infrastructure.

· Evaluate the proposal by identifying and improving the resilience indicators, based on the variables that the model has.

3 Methodology

The development methodology of the research project, based on modeling and simulation, consists of the following steps:

HED's investigation and analysis of disrupted situations. At this stage, an exhaustive analysis of the relevant bibliography will be carried out to determine the situations in which critical situations may exist and to identify performance indicators.

The Design of a simulation model in which dangerous situations are provoked to obtain an adequate response to each one of them. The model will allow the generation of simulated data for later evaluation with real data, thus forming the extended data as shown in Figure 1.

Measure recovery and obtain resilience indices: The complete data with adverse scenarios will be used to identify resilience and predict the behavior of the system in future scenarios, incorporating the SDG objectives specified in the project objective. The indices obtained will provide feedback on the prediction of events and will be fed by preventive and reactive situations in order to improve their measurement.

Obtain information that allows making decisions in emergencies according to the model. It would be achieved through preventive and reactive actions that will be public and will facilitate the generation of knowledge.

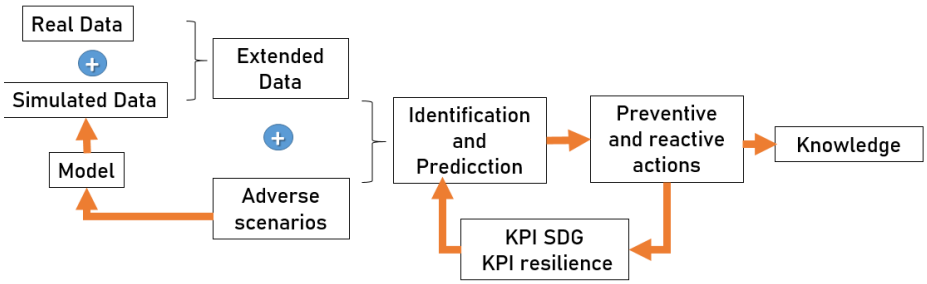


Fig. 1. Simulation process of the Hospital Emergency System in adverse scenarios.

The model to be developed must obtain as output the capacity of organization and infrastructure to support and recover from adverse situations, in order to provide and maintain its operation and services in an acceptable manner. In this way, we measure an aspect as important as availability, the capabilities available and, in the end, anticipate, resist, recover and evolve in the face of future risks and threats.

4 Conclusions

The research is part of the line of Applications with Social Impact of the High Performance Computing (HPC) research group for Efficient Applications and simulation that seeks to create technological solutions with the capacity of HPC systems. The project is in its initial research stage and aims to be an aid for decision-making in HEDs when extreme situations of reduction or overflow in patient care arise.

The next steps will be to analyze the situations of acute danger that will bring the simulator to a turning point. Finally, it is important to define indicators such as: the time of patient care, the number of exits, the threshold values of critical situations, all of them are necessary to measure the resilience of the system.

5 Acknowledgment

This research has been supported by the Agencia Estatal de Investigación (AEI), Spain and the Fondo Europeo de Desarrollo Regional (FEDER) UE, under contract PID2020-112496GB-I00

References

1. Cuarto Simposio Mundial de Investigación sobre Sistemas de Salud, realizado en Vancouver, Canadá. Noviembre de 2016, https://healthsystemsresearch.org/hsr2016/sistemas-de-salud-resilientes-en-la-region-de-las-americas/ , last accessed 2022/03/12
2. 55º Concejo Directivo: Los sistemas de salud resilientes, 68º Sesión del comité regional de la OMS para las Américas, (2016)
3. Naciones Unidas: Marco de Sendai para la Reducción del Riesgo de Desastres, 2015
4. Naciones Unidas: Programa de Las Naciones Unidas para el desarrollo de los Objetivos de Desarrollo Sostenibles, emihttps://www.undp.org/es/sustainable-development-goals, last accessed 2022/03/20 .
5. Taboada Taboada, M., Cabrera, E., Epelde, F., Iglesias, M. L., & Luque, E. (2013). Using an agent-based simulation for predicting the effects of patients derivation policies in emergency departments. Procedia Computer Science, 18, 641-650
6. Bruballa, E., Wong, A., Rexachs, D., & Luque, E. (2019). An intelligent scheduling of non-critical patients admission for emergency department. IEEE Access, 8, 9209-9220