#2 DECEMBER 2021

DaSCInews

THE QUARTERLY NEWSLETTER OF THE
ANDALUSIAN RESEARCH INSTITUTE IN DATA SCIENCE AND COMPUTATIONAL INTELLIGENCE (DASCI)

AIR-Andalusia
Boosting the AI&Robotics-based digital transformation of SMEs

Thanks to Pablo García for the design of the postcard

DaSCI NEWS is a periodical publication edited by the Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI)

Issue 2
December 2021
Published in Granada
ISSN: 2792-9132

Contact and follow us on the Internet

www.dasci.es
es.DaSCI
@sintonia_dasci
@Research_DaSCI
dasci@dasci.es
@dasci_es
/dasci-research-institute

UNIVERSIDAD DE GRANADA
Universidad de Jaén
UNIVERSIDAD DE CÓRDOBA

AIR ANDALUSIA

Brief Lecture

Federated Learning: the AI paradigm for preserving data privacy

Eugenio Martínez Cámara

Artificial Intelligence (AI) is a science and technology that has reached its maturity, and its use is increasingly evident in industry, making AI to be part of our day-to-day life and to be considered as a facilitator to achieve the SDG [1].

Data-driven methods dominate the state-of-the-art of AI, and they incessantly require more and more data. This reliance on the availability of data evidences that AI has to face up new challenges, specifically:

1. Data volume. The models from the state of the art of most of the diverse applications of AI are eager of data, and specially those models based on deep learning. Big Data techniques are able to work with huge amounts of data, but they may not be the most adequate option when the data is distributed between several data storage servers or in those cases where the data cannot be shared, because of domain requirements, communication costs or legal restrictions. Distributed machine learning emerges as a solution for the distributed processing of data, but it is neither valid for the data privacy challenge, nor for a scenario with a large number of nodes and a non homogeneous data distribution [2].

2. Data distribution. It is impractical to think that the data will be always available in one place in the AI of the near future. Accordingly, the data will be distributed among several storage devices or even in their source devices. The data from the same domain distributed in independent datasets may follow two kinds of distributions: **iid** (independent and identically distributed) and **non-iid** (non-independent and identically distributed). In scenarios where the size of nodes is large, it is much more likely that the distribution of data in each node follows a non-iid distribution, or in other words, a non homogeneous data distribution [3].

3. Data privacy. The preservation of data privacy is becoming a requirement for AI systems, since the increasing awareness of people related to the relevancy of their personal data and the presumably legal regulations on protecting data privacy stemmed from the current recommendations on trustworthy AI [4]. Regarding the previous two challenges, if the data has to be processed in a distributed way, there exists a risk of privacy leakage in case the data has to be shared among the data source and the processing node. Hence, we need a learning approach that allows jointly to process data sequestered in their data owner nodes.

4. Data integrity. AI systems have to be robust against adversarial attacks, which may impair the integrity of the learning models and the privacy of data. This requirement is stronger in distributed scenarios and when the data is not accessible because of privacy reasons.

Federated Learning (FL) has emerged as a learning paradigm to address these four challenges. FL is a distributed machine learning paradigm orchestrated by an aggregating server that collaboratively trains a learning model keeping the data in their data silos [5]. The federated learning model is trained by the iterative aggregation of learning models trained on data silos, in which the problem data is stored, and more specifically:

1. Each client trains a local model with its data.
2. Once the local models are trained, clients share them with the server. The server aggregates using an aggregation operator all the local models into one global model that would contain the information from the local models.

3. Finally, the global model is shared with all the clients, and the clients will change its local model by the global model received. This process is repeated multiple times until the global model converges.

Several elements emerge from the previous description of the training process in FL. We call those elements the key elements of FL, which most of them are shared with standard machine learning models and others are particular to FL. Specifically, those key elements are [3]:

Data. It plays a central role as in standard machine learning, since it is the fuel of the learning process. The data is kept in the clients throughout the training process.

Learning model. It is the machine learning model run in each client of the federated learning architecture.

Federated aggregation operator. The main feature of FL is the aggregation of learning models of the clients, which is conducted by an aggregation operator. The most widely used federated aggregation operator is FedAVG, which was proposed by Google in 2017 [6].

Clients.They are the data owner nodes of the FL architecture and where the learning models are trained.

Federated server. It corresponds to the node responsible for running the aggregation operator, and it thus aggregates the learning models of the clients.

Communication among the federated server and the clients. The parameters of the learning models are shared among the federated server and the clients. The communication channel has to be efficient in order to reduce the communication latency time. Besides, the federated aggregation operators have to accelerate the convergence of federated learning models with the aim of reducing the learning rounds, which means to lessen the communication rounds between the clients and the server.

The distributed nature of FL enables it to conduct new learning problems, which originates to three categories of FL [7]. The most common situation is when all the clients share the feature and label space, but they do not share the example space, or in other words, each client has its own data examples but all of them are represented with the same features. This scenario is known as **Horizontal Federated Learning**. The second category is **Vertical Federated Learning** and corresponds when the clients share the example space, but they do not share the feature and label space. It seems that this FL category does not match with any real situation, but we expose a scenario where different companies want to make a learning model on their common clients without sharing the particular information that each of them has from their clients. The third category is **Transfer Federated Learning** and it takes place in those problems where the overlapping among the feature, label and example space of the clients is minimal.

The capacity of FL to protect data privacy by allowing learning from sequestered data will enable AI to land on new domains of application. For instance, FL will boost the use of health and genomic data in AI systems, since it will allow to aggregate data from different sources, which may be located in different parts of the world, without joining them in a common data warehouse. Besides, FL will allow companies, such as banks, telecom or insurance companies, to train learning models without sharing their data, which will make them more competitive and will further them to offer better services to their customers. FL learning will facilitate AI to resolve new problems, which means that AI will provide novel solutions to keep serving society.

REFERENCES

[1] Montes, R., Melero, F.J., Palomares, I., Alonso, S., Chiachío, J., Chiachío, M., Molina, D., Martínez-Cámara, E., Tabik, S., Herrera, F. Inteligencia Artificial y Tecnologías Digitales para los ODS. Ed. Spanish Royal Academy of Engineering, January 2021. ISBN: 978-84-95662-81-1 N. Pages: 532

[2] J. Konen et al., "Federated learning: Strategies for improving communication efficiency," in NIPS Workshop on Private Multi-Party Machine Learning, 2016.

[3] Rodríguez-Barroso, N., Stipcich, G., Jiménez-López, D., Ruiz-Millán, J. A., Martínez-Cámara, E., González-Seco, G., ... & Herrera, F. (2020). Federated Learning and Differential Privacy: Software tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy. Information Fusion, 64, 270-292.

[4] European Commission. High-Level Expert Group on AI. 2019. Ethics Guidelines for Trustworthy Artificial Intelligence.

[5] McMahan, H. B. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning, 14(1).

[6] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, vol. 54, 2017, pp. 1273–1282.

[7] Yang Q., Liu Y., Cheng Y., Kang Y., Chen T., Yu H. Federated Learning, vol. 13 Morgan & Claypool (2019), pp. 1-207.

News

Digital Innovation Hub AIR-Andalusia

A novel support tool for SMEs and public administrations in the field of technology and innovation.

The EDIH AIR Andalusia candidature is coordinated by the University of Granada and has 27 partner institutions.

This consortium is constituted as a representation of the entire Andalusian ecosystem: from universities and research centres to business associations, chambers of commerce, parks, European Business and Innovation Centres (CEEIs/BICS), companies and technology clusters.

The Rector of the University of Granada, Pilar Aranda; the Regional Minister for Economic Transformation, Industry, Knowledge and Universities of the Regional Government of Andalusia, Rogelio Velasco; the Mayor of Granada, Francisco Cuenca, and representatives of the Spanish Confederation of Small and Medium Enterprises (CEPYME) presented the Andalusian Digital Innovation Centre AIR-Andalusia, together with representatives of all the entities that make up the consortium.

AIR-Andalusia will focus in applied artificial intelligence and robotics. This organisation has been created as a support tool for small and medium-sized enterprises (SMEs) and public administrations to improve their efficiency and competitiveness, providing them with access to different digitalisation services focused on the optimisation of business/production processes, products or services that use digital technologies.

The hub also aims to facilitate access to know-how and experimentation, so that companies can "test before invest". The consortium will combine the power of artificial intelligence, robotics and data as key drivers of innovation and economic growth.

Promoted by the University of Granada and the Regional Ministry of Economic Transformation, Industry, Knowledge and Universities, AIR-Andalusia represents the entire Andalusian ecosystem: from universities and research centres to business associations, chambers of commerce, parks, European Business and Innovation Centres (CEEIs/BICS), companies and technology clusters.

All the entities that make up AIR-Andalusia have it clear that the areas of knowledge of artificial intelligence and applied robotics are those with the greatest impact and relevance for the digitalisation of industry in Andalusia.

AIR-Andalusia has been recognised by the Ministry of Industry, Trade and Tourism as a 'European Digital Innovation Hub' (EDIH) candidate to participate in the European call for the Initial Network of EDIHs of the Digital Europe programme. This call will lead to recognition by the European Union.

Each EDIH will act as an access point to the European network of EDIHs, helping local companies and/or public actors to obtain support from other EDIHs in case the necessary competences fall outside its field of competence, ensuring that each stakeholder obtains the necessary support where it is available in Europe. Conversely, each EDIH will support companies and public actors in other regions and countries introduced by other EDIHs in need of their expertise.

DaSCI is a core institution in the EDIH AIR Andalusia proposal, being the largest AI research institute

AIR-Andalusia entities have the resources and capabilities to address effective and efficient research, development and innovation processes to support SMEs in the region in their digitisation processes.

The institutions that make up the consortium aim to strengthen the Andalusian industrial ecosystem to place it at the forefront of technological advances, driving and accelerating innovation in artificial intelligence and robotics in all areas of the market.

Book Spotlight

Artificial Intelligence and Digital Technologies for the SDGs

AI and digital technologies are fundamental tools to advance on the path ahead in this decade, with an ineluctable moral and ethical responsibility towards today's world. It is a great opportunity, and an enormous challenge, to progress towards achieving the 17 Sustainable Development Goals.

The DaSCI Research Institute, has examined in this book the 17 Sustainable Development Goals (SDGs) by reviewing more than a thousand references, with a view of understanding how engineering and the technology solutions --strongly anchored in Artificial Intelligence (AI)-- can help attain the goals.

The UN established the 17 Sustainable Development Goals (SDGs) to protect the planet and ensure prosperity for all. The goals signal a paradigm shift in the way companies and governments design new business models and public policies based on sustainability. Governments, the private sector and civil society all have an important role to play in this regard.

The book, entitled "Inteligencia Artificial y Tecnologías Digitales para los ODS" (Artificial Intelligence and Digital Technologies for the SDGs), is organized into three parts:

- an introduction to artificial intelligence and digital technologies,
- an analysis of their application in achieving the SDGs, and
- a set of recommendations on actions that may lead to the execution of projects and contribute to the attainment of the associated targets.

In this connection, specialist scientific literature was reviewed, including more than a thousand bibliographic references on the 169 targets that are proposed in order to achieve the SDGs.

This book is an important contribution to ascertaining the analytical capacity of engineering under the umbrella of artificial intelligence and digitalization in the service of the SDGs, and to advancing in overcoming the challenges facing the global economy and society in the 21st century. It is also helpful for understanding the three dimensions of sustainability:

1) the economic dimension (including economic and technological development and life),
2) the social dimension (including social development and equality), and
3) the environmental dimension (including resources and environment).

The book concludes with a brief discussion revolving around five key lessons learned:

1. Data is the common foundation on which AI and digital technologies are built. Unified, accessible open data are needed to implement projects that will lead to progress in many of the challenges. Governments and businesses must converge towards this goal, generating and sharing data to enable projects to be undertaken and solutions to be designed to address the SDG targets.

2. There is an urgent need to strengthen the links between science and engineering, industry and governments by engaging in dialogue and expanding avenues for quality data.

3. The SDGs set global targets, but not all the world's countries and regions are currently in the same position in the race to reach these targets. It is therefore clear that the application of AI and digital technologies must be tailored to each country's situation, and carried out in connection with the SDGs that are most urgent.

4. Digital technologies are advancing by leaps and bounds, and this also means that it is important to look for alternative ways of measuring the fulfilment of the SDGs that are adapted to this accelerated pace of progress and the emergence of new digital paradigms.

5. We must look at the current global situation created by the COVID-19 pandemic, which has indisputably had a profound impact on all dimensions of the SDGs, far beyond the strict realm of healthcare.

Technical sheet

Title Inteligencia Artificial y Tecnologías Digitales para los ODS

Authors
Rosana Montes
Francisco Javier Melero
Iván Palomares
Sergio Alonso
Juán Chiachío
Manuel Chiachío
Daniel Molina
Eugenio Martínez-Cámara
Siham Tabik
Francisco Herrera

Publisher Spain's Royal Academy of Engineering

Year 2021

Pages 532

ISBN 978-84-95662-81-1

URL https://dasci.es/outreach/ai-loves-sdg/

Paper Spotlight

CI-Dataset and DetDSCI Methodology for Detecting Too Small and Too Large Critical Infrastructures in Satellite Images: Airports and Electrical Substations as Case Study

F. Pérez-Hernández, J. Rodríguez-Ortega, Y. Benhammou, F. Herrera and S. Tabik,

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2021).

DOI: 10.1109/JSTARS.2021.3128994

The detection of critical infrastructures in large territories represented by aerial and satellite images is of high importance in several fields such as in security, anomaly detection, land use planning, and land use change detection. However, the detection of such infrastructures is complex as they have highly variable shapes and sizes, i.e., some infrastructures, such as electrical substations, are too small while others, such as airports, are too large. Besides, airports can have a surface area either small or too large with completely different shapes, which makes its correct detection challenging. As far as we know, these limitations have not been tackled yet in previous works. This article presents 1) a smart critical infrastructure (CI) dataset, named CI-dataset, organized into two scales, small and large scales critical infrastructures and 2) a two-level resolution-independent critical infrastructure detection (DetDSCI) methodology that first determines the spatial resolution of the input image using a classification model, then analyses the image using the appropriate detector for that spatial resolution. The study targets two representative classes, airports and electrical substations. Our experiments show that DetDSCI methodology achieves up to 37.53% F1 improvement with respect to Faster R-CNN, one of the most influential detection models.

Up To Date

Latest PhD Theses

May 7, 2021

Learning rules in data stream mining: algorithms and applications

Elena Ruiz Sánchez

Research in data stream mining has been mainly focused on classification and concept switching. In addition, in classification, the main focus has been on the predictive accuracy of the models, leaving aside other factors, such as readability, that affect the usefulness of the methods in real environments. The thesis presents several algorithmic approaches based on rule learning that are able to dynamically adapt to the data and provide readable knowledge about what is happening. In addition, the feasibility and usefulness of association rule extraction in data streams is studied in two real applications.

Director: Dr. **Jorge Casillas Barranquero**

Qualification: *Oustanding Cum Laude*

June 30, 2021

Plan species detection in aerial and satellite images usin deep learning

Anastassia Safanova

The main objective of this thesis is to develop robust and accurate DL models for the monitoring of different plant species using UAV images. It presents one of the first studies in exploring the potential of deep CNNs, data preprocessing and high resolution RS data, in addressing plant species conservation problems.

Directors: Dr. **Siham Tabik** and Dr. **Yuriy Maglinets**

Qualification: *Oustanding Cum Laude*

October 21, 2021

New Methods based on Soft Computing for Calibration of Agent-Based Models: Applications in Marketing and Political Science

Ignacio Moya Señas

This PhD dissertation deals with the complex problem of calibrating and validating agent-based models (ABMs). It improves the existing calibration techniques by using novel metaheuristics and proposing an integral framework for the calibration of models considering multiple key performance indicators. These advances are presented and applied to several examples modeling actual scenarios from political sciences and marketing.

Directors: Dr. **Manuel Chica** and Dr. **Óscar Cordón**

Qualification: *Oustanding Cum Laude*

September 2, 2021

Characterization of brain images using alpha-stable distributions and isosurfaces

Diego Castillo Barnés

In this thesis we have tried to improve the characterization of neurodegenerative disorders such as Alzheimer's or Parkinson's disease. For this purpose, several methods for modeling neurological data with alpha-stable distributions have been developed; the use of isosurfaces for the extraction of morphological features from functional brain images has been proposed; and a new ensemble learning methodology capable of combining all the information coming from several heterogeneous data sources including the brain image itself and genetic or proteonic factors, among others, has been designed.

Directors: Dr. **Diego Salas González** and **Javier Ramírez Pérez de Inestrosa**

Qualification: *Oustanding Cum Laude*

Interview

“Machines with cognitive capabilities will change the dynamics of work, health and leisure”

We had the opportunity to interview Dr. Marco Pedersoli after his presentation in our DaSCI’s seminars. The title of his talk was “Efficient Deep Learning” where he introduced the most common families of approaches used to reduce the requirements of DL methods in terms of memory and computation for both training and deployment, and show how a reduction of the model footprint does not always produce a corresponding speed-up.

DaSCI: Marco, the first question we wanted to ask is related to GPU speed and low energy efficient devices. Would it be right to say that GPUs are just the opposite of Green Computing? Can GPUs be “green” some day not far away?

Marco Pedersoli: You’re right, from what I know, GPUs are not energy efficient at all. But at the same time, so far, they are the only available option to do computation on lots of data. I know that there are many, many different projects that are trying to solve this problem, that is, to be much more efficient than GPUs. And also to be able to work with sparse data, because one of the drawbacks of GPUs is that they can be efficient only when they have dense data. But with sparse data it’s a big problem. So, yes, so far, we are in an initial stage where we use GPUs because it is the only thing available. So with very large models, you can end up with a very expensive cost in terms of electricity, and then in terms of footprint. But now, many companies are trying to optimize GPUs for deep learning and this optimization is in terms, not only of computation, but also in terms of reducing the ecological footprint.

Do you think that small research groups or groups with low budgets would have a hard time working with GPUs, that is, in comparison with enterprises or bigger research groups that have more hardware resources?

Yes, for sure, nowadays to be able to perform a good research we need, as always, good ideas, but we also need a lot of resources. That is one of the reasons why I chose to do research in Efficient Computing. With efficient computing, we can find methods that work with relatively low budget and low resources, thus enabling small companies or small research centers to perform deep learning and to train their own models. But at the same time, the more resources we have, the better. And for that I think it’s really important for a small center or group to associate with other groups or centers to be able to scale up, because if everyone gets busy with their own GPUs, it doesn’t really scale up. If you start to build a cluster of GPUs that could be shared among other people, the same resource can be used in a more efficient way because sometimes we have deadlines, and we have to use a lot of these resources. Some other times, other people have a deadline, and they’ll use the same resources. So, it’s important to scale up and to put these resources together to be able to compete with the big companies and their large resources. In Canada, for instance, it’s quite nice because they have what is called Compute Canada, where we have a big cluster of computers that every professor in Canada can have access to. There are two ways to access it. One option is to use the common access to everyone, where priority is based on how much you have already been using it and on how many people are currently using it. The other option it’s also even more interesting, that is, you can apply for resources. And then, if you win the grant, you will have some specific resources reserved for you and for your projects..

Marco Pedersoli

PhD in Computer Science

Assistant Professor at ETS Montreal. He obtained his PhD in computer science in 2012 at the Autonomous University of Barcelona and the Computer Vision Center of Barcelona. Then, he was a postdoctoral fellow in computer vision and machine learning at KU Leuven with Prof. Tuytelaars and later at INRIA Grenoble with Drs. Verbeek and Schmid. At ETS Montreal he is a member of LIVIA and he is co-chairing an industrial Chair on Embedded Neural Networks for Connected Building Control. His research is mostly applied to visual recognition, the automatic interpretation and understanding of images and videos. His specific focus is on reducing the complexity and the amount of annotation required for deep learning algorithms such as convolutional and recurrent neural networks.

GPUs have evolved over time, they have more and more computing power each year. It's easier to have already trained models and later prune and retrain, or just buy another big GPU and plug it in with the others? Which one do you think is the future in this field?

That's the thing about science, it depends a lot on what you actually want to do. If the aim is to use a model that has been already trained on other data and, maybe adapt it to your specific domain, approaches like pruning or distillation can be good ideas. But, if you really want to evaluate your models on a large data set, then there are also other techniques that can still reduce the computational cost of your model using better architectures. Training can be shorter even if you have to train with a lot of data. Then I will say there is no perfect solution, all of them have pros and cons.

Deep learning is going to stay. It is not a matter of fashion. Simply it works well.

Normally we would say that if your problem has a dense representation you can work with GPUs but, if you have sparse representation then GPUs are not useful. But, nowadays, there is a trend in the research of deep neural networks at a hardware level. Could then CPUs be as efficient as GPUs?

Well, if you work at hardware level with CPUs that can do XORs directly at low level (using crossbars for instance) you can obtain very good speed ups. Then a network that normally will take, let's say one second on GPU, may take the same time on CPU with these approximations... which is great. But the problem is that training at hardware level gets a bit more tricky, because it's possible, but you need to take care of some additional things. Normally, for training, they still use floating point... instead of binary representation. So, when it's about inference, you can be very fast on CPUs, but when it's about training, it's still difficult. Nevertheless, I believe it can be an interesting research topic to be able to train good binary nets, without using floating points, that will make it possible to train deep learning models directly on CPUs. But normally, the performance is a bit lower, but it's still a great advantage.

Binary nets are very useful for computer vision, but perhaps not applicable to other things, like recurrent networks. Then, how specific to a particular field are these kinds of architectures?

It depends a lot on the technique, but in general they're quite, quite general. It's basically the idea of machine learning, right? That you don't want to care too much about the specific conditions of the problem, but you want to solve a general problem that can be applied to different domains. The techniques I presented in the talk were used for computer vision, since I do research mostly in this area. They make use of convolution neural networks and I would say that most of the techniques that I explained work very well for convolution neural networks, but they can also work with any other type of neural networks. And, yes, they would not be highly affected if the specific domain is changed. Of course, it depends on the kind of data yo have. Maybe if the input data is very sparse, one technique can be better than the other, but in general, all of them should work.

When deep learning started to become fashionable, everybody was talking about it. But artificial neural networks had their moment years ago, and it's now somehow in resurgence. Are we expecting deep learning to be with us for a long time? Or it will be just a fashion and something else will replace it? What are your thoughts about this topic?

From my perspective, I think you're right, that even in research we have this kind of fashion trends. Some topics become very fashionable while others do not, and we should find ways not to do that. Because we need to be able to follow completely different directions if we really want to find new and interesting ideas. It's important to make sure that not only the fashionable topics will receive money, but also other lines of research as well. But that said, I think that deep learning is going to stay, because we have seen that it's not just about fashion, in the sense that it works well. It requires, of course, lots of data and lots of computation. But for this, there is also a lot of research trying to reduce the amount of data that we really need, and the amount of computation also. So in my opinion, it's going to stay and it's going to evolve. Because if we see what deep learning was 10 years ago, it's very different from what it is now. So I think it's going to evolve and stay in fashion for a long time.and as a society. Communication also changes radically when any person becomes a means of communication in itself. Finally, machines with cognitive capabilities will change the dynamics of work, health and leisure.

You mentioned in the talk that pruning makes networks more efficient, but you also mentioned that when using typical architectures (e.g., VGG, AlexNet) you could even delete almost 90% of all the weights keeping the same accuracy. We know that doing this will affect the efficiency, of course, and reduce the computation but... having less connections could also have another advantage: interpretability. Have you evaluated that?

Big corporations can reach levels of power comsuption that makes no sense

What you're saying makes sense, but it will always depend on the order that we are talking about. For instance, if we have an order of 10 weights or 10 neurons, we can probably check them manually. If it's on the order of 1000s or millions, it will be impossible. But yes, I haven't thought about that, it could be exploited in some way to understand a bit more what is happening inside the neural network.
If you have a good way for pruning it means that you have a probability of the importance of each feature and maybe with that, you can even select the few that have the highest probability or highest magnitude and then check them: What do they represent?

Also power consumption and carbon footprint are beginning to be reported in the literature when working with deep learning. Do you think these will be a key feature in future research papers?

Yes, I think it's important for many reasons. First, to at least give an idea and an understanding of what has been done. Also, in terms of, as you said, carbon footprint, because especially big corporations can reach certain levels of power consumption that do not make any sense. For instance, optimizing a model to be able to perform inference 1.5 times faster than before may waste millions of dollars to try all the possible configurations that lead to this improvement. So, it's important to see the full picture and not only the final result. We need to see how they got there, how much computation they had to do, because it's unfair, not only in terms of ethics, but also in terms of comparing the work of small labs with big labs.
It's a bit like when you produce some goods. It's not just about the cost of producing them, but you should also consider the cost of eliminating them. Sometimes companies don't care much about the materials that they use for certain products because then, the elimination of these products wouldn't be part of their cost, it will be part of the cost of the community. So they try to optimize and even if they use materials that pollute more, they don't care much because they care just about their profit and not the cost of the community to eliminate the pollutants. Maybe the analogy is a bit far fetched, but you get the idea.

Thank you very much for your participation in this interview, Marco. We hope your insight about the topics we discussed are as interesting to the readers as they were to us.

DaSCI Predoc Experiences

Nuria Rodríguez

Nuria, what inspired you to go into science?

I think the main factor was my interest in data science, and how complicated it is to do data science in a company nowadays. Usually, companies get stuck in methodologies that are not very cutting edge, and they call data science doing access to a database. In the last years of my degree and my master's degree, I became interested in new branches of data science such as natural language processing and federated learning, branches that are more developed in the scientific field.

Today, with a little more insight into the scientific and business world, I do not regret my decision, although I am quite clear that my professional future will be in the private industry. I believe that my time in the world of research will provide me with very valuable tools and learning in order to develop my professional career.

Natural Language Processing and Federated Learning are Nuria's topics of interest

If you could go back in time and visit you when you were in high school... What advice would you give yourself?

I would advise myself that nothing is that important, and to focus my efforts on learning about things that really interest me. I think that during high school we are so focused on the subjects we are taught that we don't stop to think about what we are really good at, or what we really like. I luckily had the opportunity to develop my interest in mathematics from high school, which brought me to where I am today, so I think it's very important to focus on the things that really bring satisfaction.

In terms of your current research, do you think industry will be reluctant to adopt new federated learning tools?

Industry may be reluctant, but the truth is that more and more companies will have to ride the federated learning wave. Data privacy policies are becoming increasingly restrictive with regard to the use of personal user information. Therefore, if companies want to continue developing artificial intelligence models on devices containing compromised or sensitive information, they will have no choice but to look for alternatives, and federated learning is a good opportunity.

I luckily had the opportunity to develop my interest in maths from high school, which brought me to where I am today.

What is your "scientific" goal for the new year?

One of the little thorns I have left from the development of my TFM is to make an original proposal in the field of natural language processing. Due to events, I ended up changing the direction of my thesis towards federated learning. However, this year I am starting to develop a proposal that combines both lines, which fortuitously match very well. I would like to develop a really valuable proposal along these combined lines.

What would be your ideal team?

At the moment, I think my ideal job would be to collaborate in a team, large enough for the exchange of ideas to flow, but small enough so that coordination is not an impediment. For example, 5-8 people seems ideal to me. Moreover, I believe that multidisciplinary work is important, and that it is from this exchange of views that great projects emerge.

If you had to choose between research and teaching, which would you choose and why?

Well, I have mixed feelings about this. On the one hand, research is what has made me choose this path, but at the same time it is what is making me lose interest in developing an academic career. Although I love research from an ideal point of view, analysing and developing cutting-edge models in a branch of data science that I like, the way research is approached nowadays, with the review processes of journals and the evaluation of curriculum vitae by weight, where many mediocre works are considered more valuable than one excellent work... makes me find more cons than pros. On the other hand, although teaching did not interest me at first, after teaching two subjects I am finding in it a personal satisfaction that I did not expect. Just as research is difficult to be satisfactory due to the time and methodologies of the journals, teaching is immediate satisfaction. The relationship with the students, managing to capture their interest and how the effort is immediately rewarded when you really get involved in a subject and you see that the students value it and are grateful for it.

Although right now I have not made any decision about what I want for my future, I hope to keep these two facets that are giving me so much wherever I am.

The current process of evaluation of the CV "by weight", (...) makes me find more 'cons' than 'pros' in the research career

I was born in Adra (Almeria), and since high school I have been interested in mathematics, participating in several olympiads from provincial to national level and managing to enter ESTALMAT, a national project for mathematical stimulation. I decided to study the Double Degree in Computer Engineering and Mathematics at the University of Granada and, later, the Master's Degree in Data Science at the same university. Since then I am still in Granada, a city that has become my home, studying the Ph.D in Federated Learning Challenges: adversarial attacks and information extraction in Natural Language Processing at DaSCI.

Project Spotlight

FedDAP Project

Federated Learning for Preserving Data Privacy

Artificial intelligence (AI) has to address new challenges in the near future stemmed from its own progress and new demands from the society. The data-driven approach currently dominates AI, and this reliance on data is evident in the unceasing growth of the size of datasets used for training the strongest AI methods from the state of the art in the diverse AI fields and tasks. Besides, the data may come from different sources and it is impractical to join data in a common data center in some scenarios. Accordingly, distributed machine learning emerges as a possible solution to address the challenge of processing data in a distributed way, but it is not the most practical solution when the size of data nodes is high. Moreover, the performance of some machine learning algorithms is degraded in a distributed setting.

Nowadays, users of AI services are more aware of the relevancy of their personal data used by those AI services. Likewise, user demand a reliable use of their data and that they are not shared with third parties. On the other hand, national and supranational regulators are publishing recommendations that expound the properties of AI services would have to have for a trustworthy use of data. In summary, these demands evidence the need of developing AI methods that preserve the privacy and the integrity of data.

The FedDAP project aims at addressing the challenges of building AI methods on huge amount of data distributed in multiple nodes, at the same time the privacy and integrity of data is preserved. Accordingly, we are going to work in furtherance the fundamental and applied research on federated learning in the FedDAP project. Federated learning is a recent proposed learning paradigm that has attracted the research community because it allows to collaboratively train a learning model by iteratively aggregating learning models trained in the data owner nodes where them are kept sequestered.

This project will contribute to the furtherance of federated learning and to prepare it for its application in real scenarios

The FedDAP project will work on the different challenges of federated learning, paying special attention to:

1. Adapting machine learning algorithms to federated learning. We are going to work in the development of new federated aggregation operators in oder to enable the use of the core machine learning algorithms. We are specially interested on the use of interpretable algorithms, as decision trees are, in order to also boost the progress of interpretability in federated learning.

2. Federated Learning categories. The nature of federated allows different kind data distributions that enable different sort of federated learning settings. These settings correspond to the federated learning categories: horizontal, vertical and transfer. We will work in FedDAP in developing federated learning methods for the three categories of federated learning.

3. Personalization. One of the outcomes of federated learning is the generation of a global learning model from the aggregation of several local learning models. Accordingly, the global model may not represent the particularities of the local datasets. In order to keep a balance among generalization and fitting to the local datasets, we will work on the development of personalization methods.

4. Protecting from adversarial attacks. The inaccessibility of the training data represents a real challenge, since the data may be used to address adversarial attacks against the federated learning model. Accordingly, we will work on the development of robust aggregation methods against adversarial attacks in order to protect the integrity of the learning model and the privacy of data.

This project will contribute to the furtherance of federated learning and to prepare it for its application in real scenarios, such as the classification of x-ray images of COVID or the processing of private textual data of users.

RESEARCHERS

Martínez Cámara, Eugenio

Luzón García, María Victoria

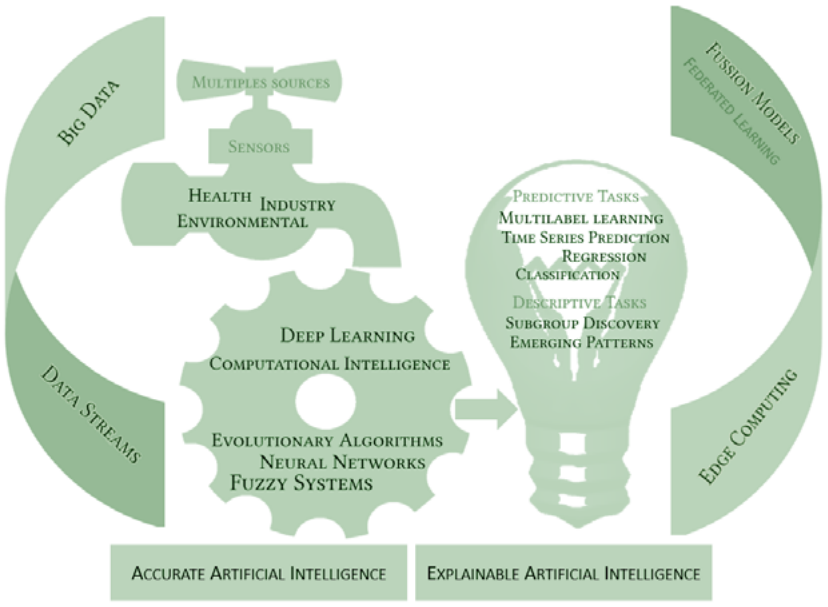
Tabik, Siham

Funding Programme	State Program for R&D&I Oriented to the Challenges of Society.
Funding Entity	State Research Agency. Ministry of Science, Innovation and Universities
Principal Researcher	Eugenio Martínez Cámara
Start	01/09/2021
End	31/08/2024
Partners	University of Granada

Project Spotlight

ToSmartEADS project

Towards a Smart, Explainable and Accurate knowledge extraction for complex Data Science problems.



Nowadays, different sources of information from Internet, social networks, sensors or other devices are gaining importance in our society. These sources generate complex data in a real time way (stream data), in large amounts, from heterogeneous sources or with multidimensional features. All these elements make up a working environment known as Big Data. Companies and institutions are interested in analysing this information to extract knowledge.

Data Science is continuously developing techniques to tackle complex problems. Within this area, Computational Intelligence stands out, where techniques such as evolutionary algorithms, systems based on fuzzy rules or neural networks are found. These include Deep Learning, where the increase in the layers of a neural network allows more precise solutions to be obtained in the problems addressed. Focusing on the characteristics of the new data sources, it is necessary to highlight those known as data stream sources, in which the generation rate does not allow its storage, so it is necessary to carry out a real time data analysis. Besides, and due to the proliferation of data sources, concepts such as information fusion are becoming very important. The information fusion at the model level, fusion of models, could lead to the discovery of new knowledge derived from the interaction between the information contained in the models. Reasons of privacy, security or data volume involves that concepts such as edge computing or federated learning appear. In edge computing, processing is as close as possible to the source of information. The concept of federated learning is related to a learning approach from distributed models. Furthermore, there is a growing interest in knowing why an Artificial Intelligence (AI) or Machine Learning (ML) system makes a certain decision. This has led to the emergence of eXplainable AI (XAI) / eXplainable ML, a research field that aims to make more understandable to humans the results and functioning of AI/ML systems.

This project aims to develop new Data Science models based on Computational Intelligence and Deep Learning to face the new challenges that emerge to extract knowledge in complex problems. This will be done from a double perspective:

a) research on new methods to obtain accurate models in the Big Data context using data streams, online processing and model fusion for classification, time series forecasting, multi-label learning, and supervised induction of descriptive rules;
b) development of transparent Data Science models for explainable AI.

The developments within the project will be accompanied by software libraries in R, Scala, Spark or Flink, that will be available to the scientific community as open source. We will address the application of the techniques developed in real problems, in the field of medicine and ecology. To this end, interdisciplinary work will be carried out in collaboration with medical researchers from the Complejo Hospitalario de Jaén and with doctors in biology from the Junta de Andalucía and WWF-ADENA.

RESEARCHERS

Charte Ojeda, Francisco
Carmona del Jesus, Cristóbal José
Gacto Colorado, María José
González García, Pedro
Martínez del Río, Francisco
Pérez Godoy, María Dolores

TEAM WORK

Álvarez Bermúdez, Antón
Garrote Alonso, Germán
López Parra, Marcos
Pérez de Ayala Balzola, Ramón
Simón Mata, Miguel Ángel
Beteta Moya , Juana
Cobo MuñoZ, José Eduardo
Fernández Herrera, David
Sáez de San Pedro Morera, Blanca
Charte Luque, David
Elizondo, David
Seker, Huseyin

Funding Programme State Program for R&D&I Oriented to the Challenges of Society.

Funding Entity State Research Agency. Ministry of Science, Innovation and Universities

Principal Researcher María José del Jesus Díaz
Antonio Jesús Rivera Rivas

Start 01/06/2020

End 31/05/2023

Partners University of Jaén

DaSCI in media

UGR Professor Oscar Cordón speaks in Almuñécar on Artificial Intelligence

The professor of the Department of Computer Science and Artificial Intelligence of the University of Granada, Oscar Cordón, will give a lecture this Friday at the Casa de la Cultura in Almuñécar entitled "Brief overview of Artificial Intelligence".

In this conference, Professor Oscar Cordón will give a historical overview of Artificial Intelligence (AI). He will outline the global, European and Spanish framework, describe those areas in which AI systems are competitive with human performance. He will also focus on explainable Artificial Intelligence, automatic systems that are capable of explaining their decisions, and on some applications of natural language processing and other areas, according to the lecturer.

COSTA DIGITAL .es


COSTA DIGITAL.es
LA REVISTA DE ALMUÑÉCAR Y LA COSTA TROPICAL

INICIO OPINIÓN ACTUALIDAD CULTURA SOCIEDAD TURIS

El catedrático de la UGR Oscar Cordón habla en Almuñécar sobre Inteligencia Artificial

Redacción Last updated 20 Nov 2021

Google ha cerrado el anuncio

El catedrático del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada, Oscar Cordón, dará una conferencia este viernes en la Casa de la Cultura de Almuñécar con el nombre de "Breve visión general de la Inteligencia Artificial"

En esta conferencia el profesor Oscar Cordón hará un recorrido histórico de la Inteligencia Artificial (IA). Planteará el marco mundial, europeo y español, describirá aquellas áreas en las que sistemas de IA son competitivos con respecto

Minsait (INDRA) and the UGR present the Artificial Intelligence Centre of Excellence 'AI Lab Granada' to the most important companies in Spain

Enrique Herrera Viedma, vice-rector for Research and Transfer at the UGR, presented the Artificial Intelligence Centre of Excellence AI Lab Granada during the inauguration of Plaiground, a collaborative tool that will open the door of AI to the business world. The event was attended by representatives from various IBEX 35 companies.

In his speech "AI Lab Granada: National strategic ecosystem in AI", Herrera Viedma highlighted the importance of AI Lab Granada in the national and European artificial intelligence strategy.
"The Centre of Excellence in Artificial Intelligence places the UGR and the city of Granada at the forefront of the AI business ecosystem in Spain and throughout Europe", said Herrera.


Granada Hoy GRANADA

Minsait (INDRA) y la UGR presentan el Centro de Excelencia en Inteligencia Artificial 'AI Lab Granada' a las empresas más importantes de España

• El vicerrector de Investigación y Transferencia, Enrique Herrera Viedma, ha dado a conocer el proyecto durante la presentación de Plaiground, la nueva herramienta que permitirá abrir el uso de la inteligencia artificial al tejido empresarial

El vicerrector de Investigación y Transferencia de la UGR, Enrique Herrera Viedma, ha presentado el Centro de Excelencia en **Inteligencia Artificial AI Lab Granada** durante la inauguración de Plaiground, una herramienta colaborativa que abrirá la puerta de la IA al tejido empresarial. A este acto han acudido representantes de diversas compañías del IBEX 35.

Granada Hoy

The UJA among the 450 best universities in the world in terms of research performance in Engineering

DIARIO JAÉN

The University of Cordoba has 39 researchers among the most influential in the world

CÓRDOBA 80 aniversario (1941-2021)

116 UGR researchers are among the most influential in the world

AULA MAGNA
El Periódico Universitario

The complete list of the most outstanding researchers at the University of Granada


Granada Hoy GRANADA

La lista completa de los investigadores más destacados de la Universidad de Granada

• La Universidad de Granada ha pasado de los 59 investigadores que tenía en el ranking de Stanford en 2020 a los 116 en 2021

Up To Date

DaSCI Webinars

The second season of DaSCI Webinars began in October 2021 with the aim of consolidating as an internal training program for all the researchers and PhD students, and as a tool to strengthen the research relationship between DaSCI members and the invited researchers.

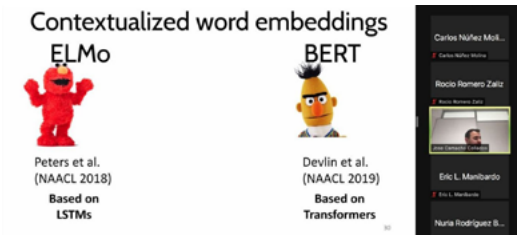
Accordingly, the DaSCI Webinar program keeps its structure of three webinar series: the Seminar serie starred by senior international researchers, the Lecture serie given by senior researchers with an established relation with DaSCI and the Reading serie performed by PhD students.

The first Webinar of the 2021-2022 season was performed by PhD Eugenio Martínez Cámara and it was about Federated Learning. This new paradigm will lead the approach of the present and future challenges of artificial intelligence, specially those ones related to the preservation of data privacy. Due to this relevancy for artificial intelligence, the second and the third webinars were also related to federated learning and the protection of data privacy.

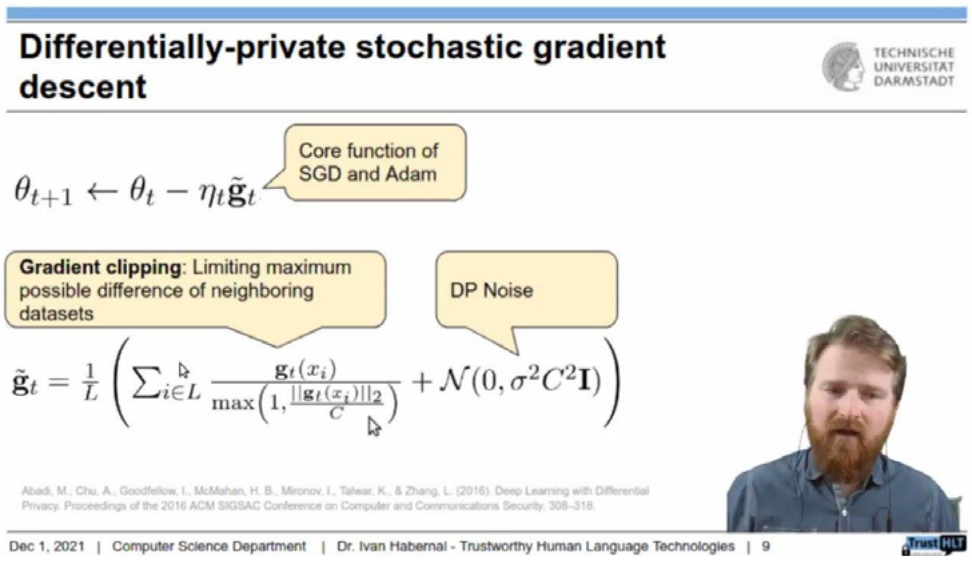
In the first days of november, Nuria Rodríguez Barroso and Daniel Jiménez López, who are two PhD students from DaSCI, presented their last advances in the development of robust federated aggregation operators against adversarial attacks for strengthening the protection of the federated model integrity and data privacy.

PhD Ivan Habernal from the Technische Universität Darmstadt (Germany), on the first day of December, illustrated us with his research line on protecting data privacy through the use of differential privacy in natural language processing tasks.

PhD Jose Camacho Collados is a senior researcher in the Cardiff University (United Kingdom), and he is specialised in researching language models, and more specially on cross-lingual language models. Currently, he is spending some months working at the University of Granada with researchers of DaSCI on how to apply federated learning in natural language processing tasks. We invited Jose Camacho to give a seminar, and on the 15th of December he exposed us to his research in word embeddings, which is framed in his book "Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning".



The DaSCI webinars will continue in 2022, and the next webinars will focus on trendy topics on artificial intelligence, for instance swarm robotics, the maths behind deep learning and autonomous driving. We will keep inviting top researchers to expound on the latest advances in hot artificial intelligence topics with the aim of keeping updated DaSCI researchers.



Federated Learning, Cibersecurity, Data Privacy or Natural Language Processing have been covered in 2021

Up To Date

Season 2 of SintonIA

@SintonIA_DaSCI - Artificial Intelligence (AI) on air.

Season 2 of SintonIA is here, and with interesting new features. With this new season, we emphasize basic and methodological dissemination, without forgetting the most striking applications of AI in fields such as astronomy or medicine.

Throughout the academic year 2021/22 we are carrying out a cycle on Artificial Neural Networks (ANN Cycle), one of the best known tools of AI. During this cycle we have talked with experts such as Jose Manuel Benítez, Anabel Gómez or Sebastián Ventura about the history of artificial neurons or the first models that existed.

To date, 4 of these episodes of the cycle on neural networks have been broadcasted, and we have even discussed advanced topics such as convolutional networks or transformers. This always without forgetting the latest developments, such as the Federated Learning paradigm, which we discuss with Nuria Rodriguez Barroso. Or our special on data science in Astronomy, Data Astroscience, in which Sergio Alonso told us about his experience collaborating with NASA. Of course, without forgetting our Christmas special... Christmas carol included!

We will be back in early 2022 with the continuation of our neural networks series, a new thematic series on bio-inspired algorithms and many more podcasts full of news and experts who will tell us about their experience and put a face to these technologies.

SintonIA episodes
Season 2

E9 *The neuron. ANN Cycle*
Guest: JOSÉ MANUEL BENÍTEZ

E10 *Data Astroscience*
Guest: SERGIO ALONSO

E11 *The perceptron. ANN Cycle*
Guest: SEBASTIÁN VENTURA

E12 *Federated Learning*
Guest: NURIA RODRÍGUEZ

E13 *Convolutional neural networks. ANN Cycle.*
Guest: ANABEL GÓMEZ

E14 *Transformers. ANN Cycle*
Guest: MARCO FORMOSO

"There is a problem we must face: the energy consumption of artificial neural networks"

JOSÉ MANUEL BENÍTEZ
SintonIA S2E9

Spreaker From iHeart

https://www.spreaker.com/show/sintonia-la-ia-en-las-ondas

Spotify

https://sl.ugr.es/SpotifySintonIA

About the DaSCI Institute

A total of 20 research projects and 12 contracts with companies partially finance the research activity of the institute, which currently has 90 PhD researchers (21.11% of whom are women) and tutors a hundred pre-doctoral students.

The communication strategy in social networks of the DaSCI institute is mainly divided into two pillars: institutional communication and science dissemination. In total, profiles on Facebook, Twitter and LinkedIn are used to provide news about the institute. A through them we work on the communication of calls for proposals and results of current projects, as well as collaborations with companies. Other profiles of the institute bring us closer to society with tiktok and instagram accounts, mainly to focus on the brand "SintonIA, La IA en las Ondas" with which our podcast is known.