ISTI NEWS

Making complex AI models practical

The widespread adoption of complex machine-learned models in critical applications asks for novel solutions aimed at making fast and efficient the off-line training of AI models and their on-line use. While the efficient off-line training of AI models is nowadays a reality due to the availability of powerful computational platforms (GPUs/TPUs) and libraries [4], how to make inference fast is still an open issue. The time needed to apply an AI model to a data instance is a critical aspect in many applications where the rate of incoming requests is high and severe constraints on prediction quality impose the adoption of complex models with millions/billions of parameters.

[page 5]

Pinocchio between art and technology

The Piazzetta dei Mosaici in Collodi comes alive thanks to augmented reality

There is an increasingly intense and evident fusion today between the concrete, material, habitable world and the virtual digital space where, at least potentially, everything is possible.

The VERO Project originates from these general premises and brings together the Carlo Collodi National Foundation and the Signal and Image Laboratory of ISTI.

[page 41]

Editorial

Taking stock of a complex year

At the end of this extremely challenging year, my first complete year as Director of ISTI, it's finally time to take stock. An unprecedented event, Covid-19 (the only global pandemic in most living memories) has characterized 2020.

[page 3]

PRAMA

Common clinical trials for Alzheimer's disease (AD) rely on outdated hypotheses on disease pathogenesis and on approximate criteria for patient selection, grouping together patients with diverse manifestations of the disease.

[page 12]

National Research Council of Italy

Contents

Taking stock of a complex year

At the end of this extremely challenging year, my first complete year as Director of ISTI, it's finally time to take stock. An unprecedented event, Covid-19 (the only global pandemic in most living memories) has characterized 2020. It is not the first time a major disease has encompassed the world, but this one has affected what is now a strongly interconnected and globalized society.

The end results have been terrible and terrifying on both social and economic levels. We discussed the effect of Covid-19 on our Institute, and the steps we took to address it in the Editorial of the previous issue of ISTI News, and I will not repeat the comments made then. The difference is that last summer, we were fairly confident that Covid-19 would gradually reduce its impact; unfortunately that has not yet happened and public health is still at great risk.

However, although it's true that Covid-19 has affected the activities of ISTI, if we examine some indicators relevant to scientific results and dissemination, our key activities, it appears that so far, the outcome is not too bad. It seems that we are more resilient than we realized. Luckily, our particular efforts have been relatively easy to transform into mainly remote, digital processes.

Scientific production. When taking stock of 2020, it is mandatory to start with an analysis of our scientific production. The end of the year is usually not the best moment for statistics of this type since publications are finalized in the last few weeks (and in some cases, in the first weeks of the new year). And I have to admit that our staff is not particularly fast in entering the relevant data in our publications archive. Thus, I present here for comparison the final figures for publications 2018-2019 and the current figures for 2020 (to be consolidated in March 2021) [Tab. 1].

Hopefully, the final data will confirm that the number of journal publications this year is similar to that of 2018 and 2019. In 2019 we produced an average of 1,35 journal papers per researcher. The distribution of journal papers per quartile, according to the classification proposed by SCOPUS, is presented in the two graphs below. It is worth mentioning that in 2019 61% of ISTI papers were published in Q1 venues [Pie chart 1]; the 2020 preliminary data evidence some decrease, but the percentage of Q1 papers is still very high (50%) [Pie chart 2]. This is an indicator of the quality of the results produced.

Conversely, 2020 shows a steep decrease in the number of conference papers. This is probably due to the pandemic and uncertainty concerning the organization of conferences in 2020. The plausible trend in the second half of 2020 was to submit research results to journals rather than conferences (thus, these papers are now under a slower review process and some at least should appear in 2021). [Pie chart 1 and 2]

Project acceptance. A second key indicator of the institute's health is success in securing funds on competitive project calls. The CNR headquarters in Rome is responsible for the salaries of tenured staff and part of the ordinary costs (we call these centralized funds). All other expenses (temporary staff, equipment, travel, doctorial grants, a share of ordinary costs, etc.) are covered by external funds. Therefore, being able to secure externally funded projects is of primary importance for us.

I am proud of the results obtained in 2020, in a context of ever-increasing competition with respect to EC and National Calls for project proposals. As you can see from the table below, ISTI researchers were awarded **19** new EC projects and a number of others at the national level. The share of EC vs. national funding also increased in 2020. In 2020, we obtained a total of 8,2 MEuro.

The external funds secured by ISTI were approximately 70-75% of the centralized funding [Tab. 2].

ISTI Day. The annual event of ISTI in which we disseminate our scientific results took place in November 2020. Due to Covid-19 measures, we decided to hold this event via teleconference in three half-day sessions, with short talks presenting recent scientific results. Videos of these presentations are available on our website (https://www.isti.cnr.it/en/research/isti-day, in Italian).

Press reports. During 2020 we had a good coverage by the general press. More than thirty articles discussing ISTI scientific results were published by different press media (national, local and international). This represents the first results of a new policy aimed at producing material targeted at the popular press, that we plan to reinforce in 2021.

In addition to these results, obviously the day-to-day running of the Institute continued, albeit in unusual circumstances. This turned out more critical for our administrative staff. Luckily our staff was not only to conduct much of its business in isolation, at home, but tools such as Zoom, WhatsApp and Skype meant that contacts and communication were facilitated and used regularly.

Some changes in ISTI organization. In the last two years, half of our research laboratories have appointed new heads, the process has been smooth and seamless.

One of the research labs, NEtworked Multimedia Information Systems (NEMIS), decided to split into two independent research labs to better reflect diverging focuses:

- **Artificial Intelligence for Media and Humanities** (AIMH), whose mission is to investigate and advance the state of Artificial Intelligence, specifically addressing applications to digital media and Digital Humanities, with attention to issues related to the scalability of technology.
- **Infrastructures for Science** (InfraScience), which aims to design and deliver

digital infrastructures, information systems, and AI-assisted solutions; scientific and technological activities that support the evolution of scientific approaches, nowadays increasingly open and collaborative, multidisciplinary, and data-intensive.

The ISTI web has been redesigned and restructured with some new material and, hopefully, easier access to the content (https://www.isti.cnr.it/en/).

History of ISTI. History is important since what we are today depends in many cases on choices or actions of many years ago. It is thus vital to keep alive the memory of our past. Therefore, we have collected together some old documents representing the history of ISTI and its predecessors (CNUCE and IEI). This material will be available soon on our web at https://www.isti.cnr.it/it/chi-siamo (temporarily just in Italian).

Let me just conclude by saying that despite

what has perhaps been the most difficult year in our Institute's lifetime, overall, I am reasonably satisfied with the results and extremely proud of our staff who have managed to keep working, mainly cheerfully, under such difficult and, at times, stressful circumstances.

Contact: Roberto Scopigno, Director
direttore@isti.cnr.it

ISTI Publications	2018	2019	2020
Journal papers	133	135	127
Conference papers	126	170	89

Tab. 1

New Projects (accepted in 2020)		
EC	19	6.377.022,00 €
National	8	453.399,00 €
Regional & Local	2	392.00,00 €
Industrial Contracts	35	1.044.385,00 €
Total	**64**	**8.266.806,00 €**

Tab. 2
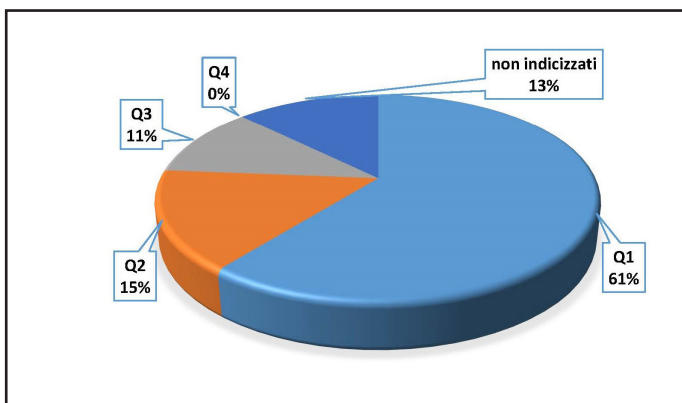
Pie chart 1 - Journal papers and Scopus quartiles 2019
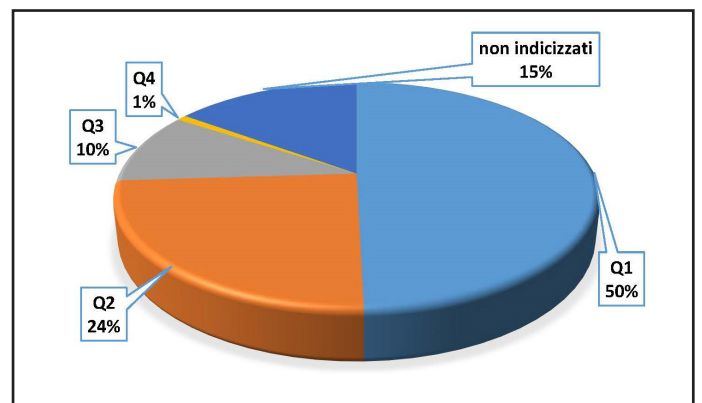
Pie chart 2 - Journal papers and Scopus quartiles 2020

Making complex AI models practical

Franco Maria Nardini, Raffaele Perego (HPC Lab, ISTI-CNR)

The widespread adoption of complex machine-learned models in critical applications means that there is now strong demand for innovative solutions aimed at speeding up the efficient off-line training of AI models and their on-line use. While the efficient off-line training of AI models is nowadays a reality due to the availability of powerful computational platforms (GPUs/TPUs) and libraries [4], how to render inference fast is still an open issue. The time needed to apply an AI model to a data instance is a critical aspect in many applications, where the rate of incoming requests is high and severe constraints on prediction quality impose the adoption of complex models with millions/billions of parameters. In these cases, quality-of-service requirements entail the optimization of AI models, subject to performing inference in near real-time or within a limited time budget. Devising novel solutions to speed up inference allows more accurate models to be deployed without violating the time constraints.

We have identified two main scenarios where finding the best trade-off between model accuracy and inference time is both important and challenging. The first scenario regards large-scale Information Retrieval systems and involves the ranking of documents according to their relevance for user queries. As is known, ranking involves challenging effectiveness and efficiency constraints in many online services deployed in Web search engines, e-Commerce platforms, online social networks, or conversational assistant systems. The second scenario refers to Industry 4.0 and the real-time analysis of (semi-finished) products on an assembly line for the identification of defects and the classification of their severity. In both these use cases, any solution advancing the state of the art will have a relevant economic impact. In this brief article, we focus specifically on the first scenario, reviewing some research results investigating AI model efficiency obtained at the HPC Lab of ISTI-CNR. We conclude discussing some open issues and future research directions.

Fast and accurate ranking in information retrieval

Ranking is a central task in many information retrieval problems, where documents must be ranked according to their relevance to an implicit or explicit query. From an effectiveness point of view, a number of AI solutions, falling under the broad category of Learning-to-Rank algorithms, have been proposed to learn from supervised data effective ranking functions. The most effective models generated by such methods are, however, in many cases too expensive to use in large-scale on-line systems. For example, state-of-the-art ranking methods based on forests of regression trees involve tens of thousands of decision trees to be evaluated for each document potentially answering the query. Current neural rankers, based on deep learning approaches, are even more computationally demanding. In particular, those making use of pre-trained contextualized language models promise to be disruptively effective, but they introduce a substantial computational burden in the ranking task. Therefore, even if AI-powered rankers are in principle highly effective, it is not possible to apply them when query rate is high and inference time budgeted.

To partially overcome this issue, large-scale information retrieval systems exploit a multi-stage ranking architecture like the one sketched in Figure 1 for *top-K* document retrieval. Here, a simple and fast *base ranker*, optimized for recall, retrieves N, $N>>K$, documents matching the query. In the second step, a complex AI-powered model is used by the *top ranker* to re-rank the candidate documents coming from the first step. The top ranker is optimized for precision, i.e., to place the most relevant results in the top positions of the ranked list of results. It constitutes a performance bottleneck for the whole rank-
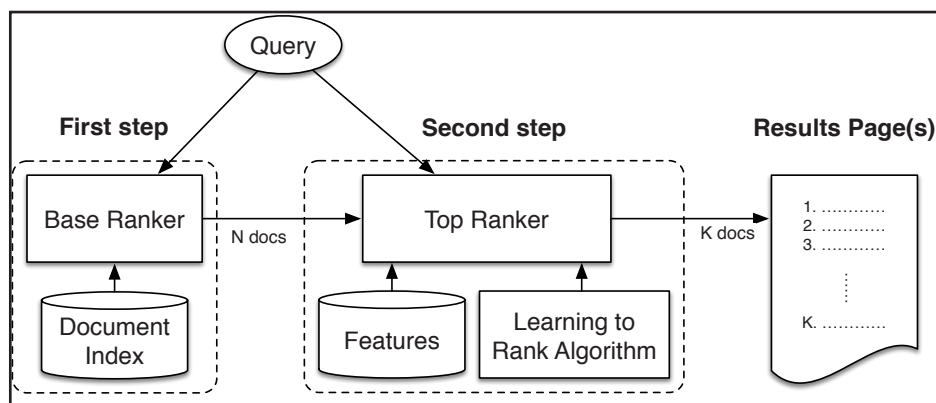

Figure 1: a two-stage ranking architecture

ing pipeline. In the past, the HPC Lab has contributed to removing this bottleneck for AI models based on large forests of decision trees [5]. Recently, we have also focused on neural rankers based on deep learning methods. In the following, we introduce a contribution aimed at making neural ranking models not only efficient but also practical.

Making neural ranking practical

Pre-trained deep transformer networks, e.g., BERT [1], have been transformative for many tasks, exceeding the prior art in several natural language processing and information retrieval tasks. Neural approaches for ranking can generally be characterized as either *representation-based or interaction-based*. Representation-based models build representations of a query and document independently and then combine these representations to calculate a relevance score. These methods are beneficial because they allow us to compute document representations at index time to reduce the latency of query processing. On the other hand, the lack of attention mechanisms modeling the interaction between query and documents may result in a lower effectiveness. The HPC lab contributed to representation-based methods with EPIC [3], a new approach for passage retrieval that computes term importance (i.e., salience) and expansion over

a contextualized language model to build interpretable query and document representations. EPIC advances the state of the art by approaching fully BERT ranking performance while providing low query-time latency and full interpretability of representations.

Instead, interaction-based models combine signals from the query and the document at query time to compute the relevance score. These models are the most effective but also the most expensive to use at run time. For example, the use of a fine-tuned BERT-based model as top ranker requires up to 3.5 seconds to re-rank the candidate documents for a single query. However, in the last two years or so, transformer networks have grown from about 110 million parameters (BERT, 2018) to 175 billion parameters (GPT-3, 2020). These network models have a huge impact on computational performance when applied to tasks like ad-hoc retrieval to the point of being, in most cases, impractical. In order to make interaction-based models practical, we proposed PreTTR (Precomputing Transformer Term Representations) [2]. This method moves from the observation that much of the term interaction at query time happens locally within either the query or document, and only the last few layers of a transformer network are required to produce effective ranking scores once these representations are built. Thus, documents can

be pre-processed at indexing time through part of the network. The output of this partial network computation is a sequence of contextualised term representations (Figure 2).

These representations can then be stored and used at query time to finish the processing in conjunction with the query (see Figure 2). At training time, the transformer network is fine-tuned for ad-hoc document ranking. This transformer network masks attention scores in the first l layers, disallowing interactions between the query and document. At index time, each document in the collection is processed through the first l layers, and the resulting term representations (green segments) are stored on disk. Finally, at query time, the query is processed through the first l layers (orange segments), and then combined with the document term representations to finish the ranking score calculation (blue segments). Since the representation of each layer can be large (e.g., 768 floating point values per document term in the base version of BERT), we also propose a compression approach to reduce the possible storage burden. This approach involves training an encoding layer between two transformer layers that produces representations that can replicate the attention patterns exhibited by the original model. We experimentally show that all these processes result in a much faster ranking at query time (up to 42x speedup
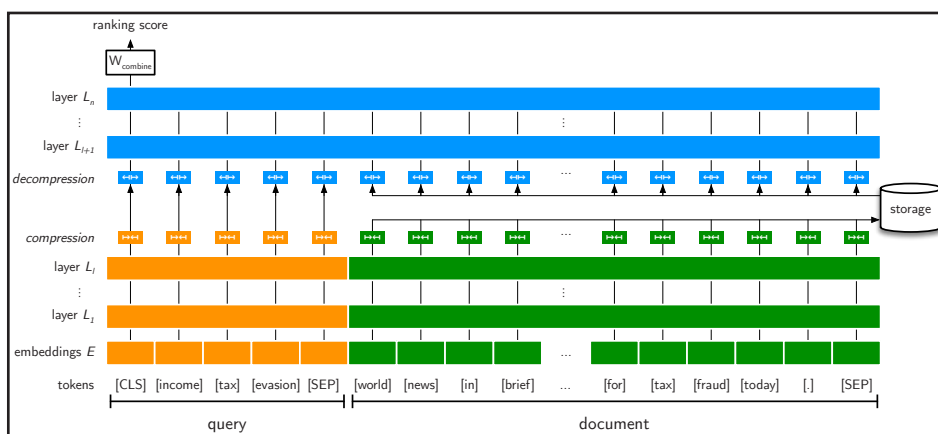

Figure 2: overview of PreTTR

over the original model), while having only a minimal impact on the ranking performance and a reasonable impact on the index size. These are all critical findings and allow transformer networks to be used practically for large-scale ranking tasks, yielding substantial improvements to ranking accuracy.

Open Issues and Challenges

Nowadays, deep neural networks are a key component in intelligent applications involving text, images, video, and combinations thereof. The astonishing effectiveness of modern neural networks is, however, achieved by increasing their complexity. This poses several challenges when using these models online, especially in applications characterized by tight inference time budgets. We show how it is possible to target

inference efficiency in an exemplary use case from information retrieval. There are, however, a number of open challenges for this line of research.

First, an investigation of to what extent the approaches proposed can be generalized to other tasks in the same or relevant fields. A second research question regards how these approaches can be combined with those dealing with over-parametrization of deep neural networks through model compression/distillation.

AI efficient inference is a topic attracting increasing interest and importance in other domains such as Industry 4.0, IoT, and fog/edge computing. It is a cross-cutting research area that requires excellent algorithmic skills and a very good knowledge of the requirements

of the specific application domain in order to optimize the use of complex AI models.

Contact: francomaria.nardini@isti.cnr.it
raffaele.perego@isti.cnr.it

References

1.   J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT. 2019.

2.   S. MacAvaney, F.M. Nardini, R. Perego, N. Tonellotto, N. Goharian, and O. Frieder. Efficient Document Re-Ranking for Transformers by Precomputing Term Representations. Proceedings of ACM SIGIR. 2020.

3.   S. MacAvaney, F.M. Nardini, R. Perego, N. Tonellotto, N. Goharian, and O. Frieder. Expansion via Prediction of Importance with Contextualization. Proceedings of ACM SIGIR. 2020.

4.   A. Paszke et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Proceedings of Neural Information Processing Systems (NeurIPS). 2019.

5.   D. Dato, C. Lucchese, F.M. Nardini, S. Orlando, R. Perego, N. Tonellotto, R. Venturini: Fast Ranking with Additive Ensembles of Oblivious and Non-Oblivious Regression Trees. ACM Trans. Inf. Syst. 35(2). 2016.

NAUTILOS

New Approach to Underwater Technologies for Innovative, Low-cost Ocean obServation Funded by HORIZON 2020

NAUTILOS – a New Approach to Underwater Technologies for Innovative, Low-cost Ocean observation is an H2020 project funded under the Future of Seas and Oceans Flagship Initiative, coordinated by the National Research Council of Italy (CNR), which brings together a group of 21 partners from 11 European countries with multidisciplinary expertise.

NAUTILOS will fill in existing marine observation and modelling gaps through the development of a new generation of cost-effective sensors and samplers for physical (salinity, temperature), chemical (inorganic carbon, nutrients, oxygen), and biological (phytoplankton, zooplankton, marine mammals) essential ocean variables.

In addition, micro-/nano-plastics sensors will be developed to improve our understanding of environmental change and anthropogenic impacts related to aquaculture, fisheries, and marine litter, as well as supporting the EU's efforts towards a European Strategy for Plastics in a Circular Economy. Newly developed marine technologies will be integrated with different observational platforms and deployed through the use of novel approaches in a broad range of crucial environmental settings (e.g. from shore to deep-sea deployments) and EU policy-relevant applications.

The fundamental aim of the project will be to complement and expand current European observational tools and services, obtaining a collection of data at a greater spatial resolution, temporal regularity and size than that currently available on European scale. A further objective will be to enable and democratise the monitoring of the marine environment for both traditional and non-traditional data users.

The NAUTILOS project will develop, integrate, validate and demonstrate new cutting-edge technologies involving sensors, interoperability and embedding skills. The goals of the project will always guide the development, focusing on the scalability, modularity, cost-effectiveness and open-source availability of the software and data. NAUTILOS will also provide full and open data feed towards well-established portals and data integrators (e.g. EMODnet[1], CMEMS[2], JERICO[3] …).

ISTI-CNR is the Coordinator of the project and will deal with Data Management aspects, from data quality assessment, to data use and sharing on well-established external repositories. Image analysis algorithms will be developed for the categorization or recognition of already available and newly acquired marine environmental images. The final goal will be to provide tools to automatically perform these tasks on different kind of images.

Contact: Gabriele Pieri, SI Lab
gabriele.pieri@isti.cnr.it
http://www.nautilos-h2020.eu

1   https://emodnet.eu/en
2   https://marine.copernicus.eu/
3   https://www.jerico-ri.eu/

HumanE-AI-Net

HumanE AI Network
Funded by HORIZON 2020

The HumanE AI Network will leverage the synergies between the involved centers of excellence to develop the scientific foundations and technological breakthroughs needed to shape the AI revolution in a direction that is beneficial to humans both individually and societally, and that adheres to European ethical values and social, cultural, legal, and political norms. The core challenge is the development of AI systems capable of "understanding" humans, adapting to complex real-world environments, and appropriately interacting in complex social settings. The aim is to facilitate AI systems that enhance human capabilities and empower individuals and society as a whole while respecting human autonomy and self-determination. The following are specific research areas will be addressed in the course of the project:

1. Human-in-the-loop machine learning, reasoning, and planning

Allowing humans to not just understand and follow the learning, reasoning, and planning process of AI systems (being explainable and accountable), but also to seamlessly interact with it with uniquely human capabilities, knowledge about the world, and the specific user's personal perspective.

2. Multimodal perception and modeling

Enabling AI systems to perceive and interpret complex real-world environments, human actions, and interactions situated in such environments and the related emotions, motivations, and social structures.

3. Human-AI collaboration and interaction.

Developing paradigms that allow humans and complex AI systems (including robotic systems and AI-enhanced environments) to interact and collaborate in a way that facilitates synergistic co-working, co-creation and enhancing each other's capabilities.

4. Societal awareness

Being able to model and understand the consequences of complex network effects in large scale mixed communities of humans and AI systems interacting over various temporal and spatial scales. This includes the ability to balance requirements related to individual users and the common good and societal concerns.

5. Legal and ethical bases for responsible AI

Ensuring that the design and use of AI is aligned with ethical principles and human values, taking into account cultural and societal context, while enabling human users to act ethically and respecting their autonomy and self-determination. This also implies that AI systems must be "under the Rule of Law": their research design, operations and output should be contestable by those affected by their decisions, and a liability for those who put them on the market.

Contact: Fosca Giannotti, KDD Lab
fosca.giannotti@isti.cnr.it
https://www.humane-ai.eu/
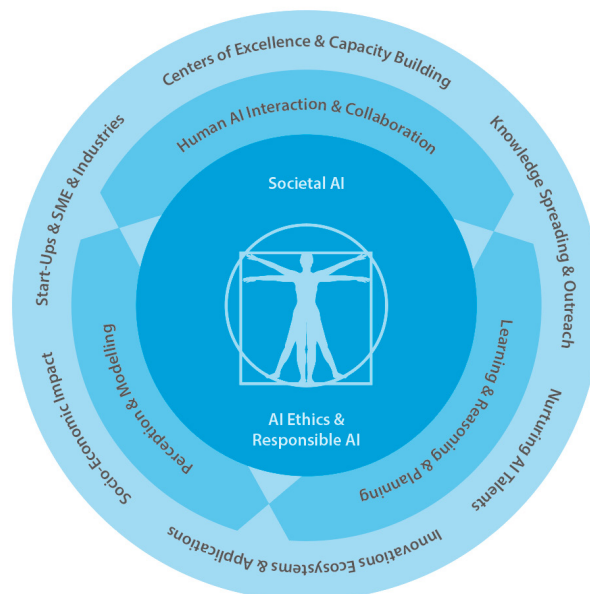
BIECO

Building trust in Ecosystems and Ecosystem Components
Funded by HORIZON 2020

Nowadays, most company-proposed ICT solutions require integration or collaboration with other ICT components, which are typically developed by third parties. Even though this practice is key in order to maintain productivity and competitiveness, the fragmentation of the supply chain can pose a high security risk, as in most cases there is no way to verify whether these components have been built in accordance with the best security practices or if they have vulnerabilities. n such situations, it is important that companies assume an "untrusted by default" position. According to a recent study, only 29% of IT businesses are sure that their ecosystem partners are compliant and resilient with regard to security. Cybersecurity attacks can have a high economic impact and it is not sufficient to rely only on trust. ICT components must provide verifiable guarantees regarding security and privacy. It is imperative that vulnerabilities in such components can be detected accurately and their ability to propagate over the supply chain and impact on ICT ecosystems must be understood. However, as most vulnerabilities can remain undetected for years and cybersecurity incidents will happen, it is necessary to provide tools that guarantee both resilience and better mitigation strategies. The horizons of the current risk assessment and auditing processes must be expanded, taking into account a much wider risk landscape.

BIECO is a holistic framework that will provide the necessary mechanisms to help companies understand and manage the cybersecurity risks and threats they are liable to when they become part of the ICT supply chain. The BIECO framework consists of a set of tools and methodologies that will address the challenges of vulnerability management, resilience, and auditing in complex systems.

Contact: Eda Marchetti, SEDC Lab
eda.marchetti@isti.cnr.it
http://www.bieco.org

TAILOR

Foundations of Trustworthy AI - Integrating reasoning, learning and optimization
Funded by HORIZON 202

TAILOR is one of four AI networks in the H2020 program ICT-48 Towards a vibrant European network of AI excellence centres.

The purpose of TAILOR is to build the capacity to provide the scientific foundations for Trustworthy AI in Europe by developing a network of centres of excellence that will leverage and combine learning, optimization and reasoning with methods for achieving explainability, safety, fairness, accountability, and sustainability.

In TAILOR, learning means Machine Learning techniques, starting with the highly successful and now prominent Deep Learning, but also including other approaches like Bayesian Networks, Support Vector Machines, and Decision Trees; optimization means continuous, discrete and combinatorial optimization, and hybrids, and includes techniques like convex optimization and gradient methods, constraint programming, heuristics and meta-heuristics, deterministic and stochastic approaches; and reasoning subsumes deductive, inductive, and abductive reasoning, monotonic and non-monotonic reasoning, and includes techniques like statistical relational learning, as well as probabilistic and inductive programming. However, to be considered trustworthy, an AI system needs to satisfy a number of requirements: respect the rule of law; be aligned with agreed ethical principles and values, including privacy, fairness, human dignity; keep the humans in control; ensure the system's behaviour is transparent, and its decision making process is explainable; and be robust and safe, meaning that the system's behaviour remains trustworthy even if things go wrong. Thus, in TAILOR, trustworthy means that all the above dimensions will be intertwined with the foundation themes through a continuous mutual exchange of requirements, challenges, and innovations.

The TAILOR project, coordinated by Linkoping University, brings together 45 of the leading AI research centres as partners and at least 60 network members from learning, optimisation, reasoning and trustworthy AI together with 10 internationally recognized European companies representing important industrial sectors into a single scientific network, guided by a strategic research and innovation roadmap thereby reducing fragmentation, boosting collaboration, increasing the impact of funding, and expanding European AI research capacity as well as developing, attracting and retaining talent in Europe.

Contact: Fosca Giannotti, KDD Lab
fosca.giannotti@isti.cnr.it

ProCAncer-I

An AI platform integrating imaging data and models, supporting precision care through prostate cancer's continuum
Funded by HORIZON 2020

Prostate cancer (PCa) is the second most frequent type of cancer in men and the third most lethal in Europe. Current clinical practices suffer from lack of precision, often leading to overdiagnosis and overtreatment of indolent tumours. This calls for advanced AI models to go beyond the current state of the art by deciphering non-intuitive, high-level medical image patterns and thus increasing performance in discriminating indolent from aggressive disease, by early predicting metastases or their recurrence, and by predicting the effectiveness of therapies. Actual efforts in the field are fragmented, based on single–institution, size-limited and vendor-specific datasets, thus making model generalizability impossible.

The ProCAncer-I project brings together 20 partners, including PCa centres of reference, world leaders in AI and innovative SMEs, with recognized expertise in their respective domains, working to design, develop and sustain a cloud based, secure **European Image Infrastructure** with tools and services for data handling. The platform will host the largest collection of PCa multi-parametric MRI, anonymized image data worldwide (>17,000 cases), in line with EU legislation through data donorship. Robust AI models will be developed, based on novel ensemble learning methodologies, leading to vendor-specific and vendor-neutral AI models for addressing eight PCa clinical scenarios. https://www.jerico-ri.eu/

To accelerate clinical translation of PCa AI models, the project will focus on improving the trust of the solutions with respect to safety, accuracy and reproducibility. Metrics to monitor model performance and inner causal relationships will cast light on model outcomes, helping to inform decision makers with respect to possible failures and errors. A roadmap for AI model certification will be defined, by interacting with regulatory authorities, thus contributing to a European regulatory roadmap for validating the effectiveness of AI-based models in clinical decision making.

ISTI-CNR's role in the project is key as the team involved is studying the development of robust AI models able to cope with the heterogeneity of imaging data and the biases and confounders this might introduce into the learning models. The team will lead the task related to AI trustworthiness, based on safety, transparency and reproducibility of results as well as on performance monitoring when used in clinical practice. Sara Colantonio will also serve as the Quality Manager of the project activities.

Contact: Contact: Sara Colantonio, SI Lab
sara.colantonio@isti.cnr.it
https://cordis.europa.eu/project/id/952159
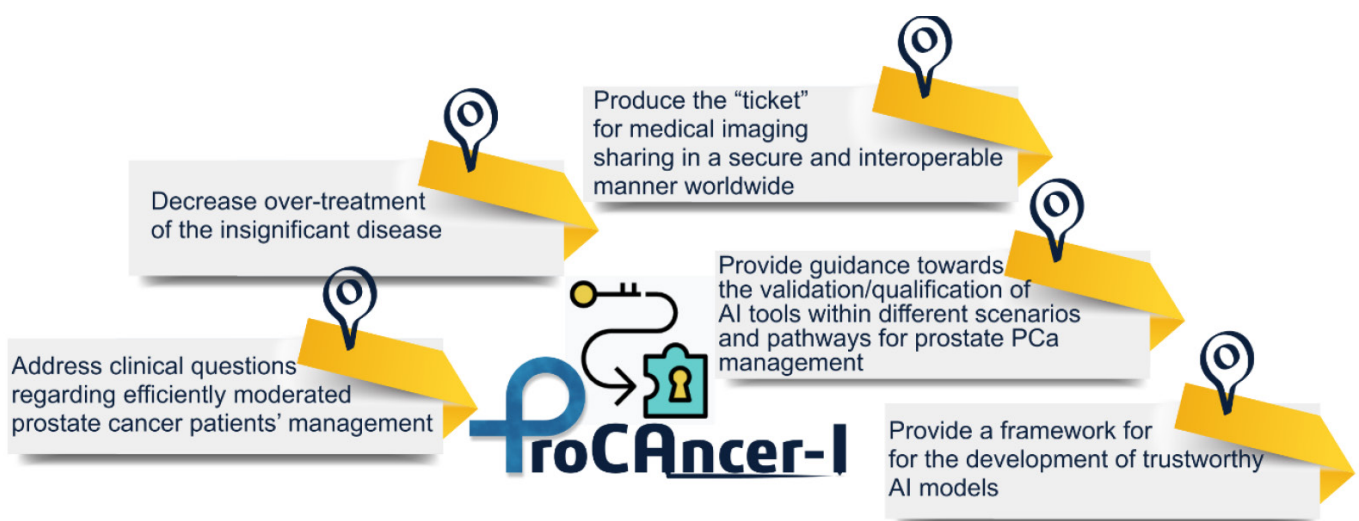
MOVING

Mountain Valorization through Interconnectedness and Green Growth
Funded by HORIZON 2020

Mountainous areas cover 36% of Europe and play an essential role in the provision of public and private welfare and goods. Despite their relevance in both ecological and socio-economic terms, updated and comparable knowledge of the impacts of climate change, demographic trends and socio-economic drivers in these areas is lacking; as are the necessary strategies to address them. Advanced knowledge on the vulnerability of land use, production systems and related value chains is needed.

Studies of the long-term trends and dynamics affecting European mountainous zones are necessary in order to be able to design the next generation of policies. The objective of MOVING is to build capacities and co-develop policy frameworks across Europe for the establishment of new or upgraded/upscaled value chains contributing to the resilience and sustainability of mountain areas, valorising local assets and delivering private and public goods. The project includes: the screening of traditional and emerging value chains in all European mountainous areas (including non-EU countries); in-depth assessment of the vulnerability and resilience of land use, production systems and value chains in 23 reference regions, taking into account the diversity of different mountainous areas; participatory theory building to provide knowledge and assessment indicators; participatory foresight exercises to provide information on future trends at regional and EU levels; benchmarking analysis to identify the drivers of the resilience of mountain ecosystems.

A policy roadmap will provide guidelines to support the design of public and private policy instruments that boost the resilience of mountainous areas and the provision of public and private goods. MOVING strongly relies on a Citizen-science-policy interface organised around a multi-actor approach: a Community of Practice formed by 23 multi-actor platforms and an EU platform; a Virtual Research Environment powered by the D4Science infrastructure (www.d4science.org) to foster online interactions among actors; and new visual tools to make the information accessible to different audiences.

Contact: Pasquale Pagano, NeMIS Lab
pasquale.pagano@isti.cnr.it

PRAMA

Proteomics, RAdiomics & Machine learning-integrated strategy for precision medicine for Alzheimer's
Co-funded by the Tuscany Region under the Par Fas Salute Toscana 2014-2020 program

Common clinical trials for Alzheimer's disease (AD) rely on outdated hypotheses on disease pathogenesis and on approximate criteria for patient selection, grouping together patients with diverse manifestations of the disease. Recent studies have suggested that AD may come with several clinical phenotypes and that the differentiation between disease subtypes can be due to the pathway followed by the AD precursor beta-amyloid (Aß) peptide when it self-assembles into amyloid aggregates in the brain. An integrated survey taking advantage of multiple marker modalities is, thus, perceived as a desirable solution to support clinicians in identifying different disease subtypes, even in their early stages, and to accordingly decide on personalized treatments for individual patients. In the PRAMA project, we intend to build up a strategy for personalized prediction of the disease based on the hypothesis that the main precursors of AD can form specific aggregates responsible for distinct clinical pictures of the disease, with consequent different sensitivity to drugs. In detail, a combined biochemical, biophysical and optical spectroscopy characterization of molecular biomarkers found in the cerebrospinal fluid of 100 individuals will be carried out, by including patients with progressive clinical signs of AD. This data will provide information on biomarker composition, structure, aggregation level and toxicity. This will constitute the proteomic profile of the biomarker content for each individual. The same patients will be subjected to magnetic resonance imaging (MRI) followed by a radiomics-based image analysis. The entire set of biochemical, optical, MRI data including clinical parameters and neuropsychological evaluation of patients will be elaborated through data analytics techniques to, firstly, discover correlations among novel and gold-standard biomarkers and, then, to mine and identify different AD phenotypes. The most recent Artificial Intelligence and Machine Learning techniques will be employed to model and process the complex high-dimensional data gathered in PRAMA. Data analyses will also aim at discovering specific diagnostic, prognostic or predictive responses at the different stages of disease stages, on a personalized basis. The outcomes of PRAMA are expected to have a high socio-economic impact, with significant advantages that include reducing healthcare costs and improving general well-being.

Contact: Sara Colantonio, SI Lab
sara.colantonio@isti.cnr.it

PlatformUptake.eu

Assessing the state of the art and supporting an evidence-based uptake and evolution of open service platforms in the Active and Healthy Ageing domain
Co-funded by Horizon 2020

PlatformUptake.eu seeks to deliver an inventory of the state of the art and analyse the use of open service platforms in the Active and Healthy Ageing population. The intention is to cover both open platforms - such as UniversAAL, FIWARE and partly-open/proprietary platforms developed by industry, also addressing the interactions between these platforms. In order to measure the impact of these platforms and en-

hance their uptake, the project will adopt a methodology for monitoring open platform development, adoption and spread across Europe. The key factors that determine success or hinderance in their uptake by end-user groups will be studied, and the evolution of their ecosystems and stakeholder networks will be investigated.

The project will assess the societal impact of existing platforms in the AHA domain, create monitoring and evaluation toolkits, collect successful user stories and best practices, promote interoperability and define guidelines for their common evolution within existing policy frameworks and initiatives. With the aim of supporting the large-scale uptake of these platforms, the project

proposes the creation of an online information hub which will provide descriptive and support materials for all existing platforms. PlatformUptake will organize a number of stakeholder events as well as an Open Online Course to promote synergies, knowledge exchange and a common understanding among all stakeholders in the Active and Healthy Ageing market.

Contact: Andrea Carboni, SI Lab,
Dario Russo, WN Lab
andrea.carboni@isti.cnr.it
dario.russo@isti.cnr.it
http://www.platformuptake.eu

AI4Media

A European Excellence Centre for Media, Society and Democracy
Funded by HORIZON 2020

Motivated by the challenges, risks and opportunities that the increasing use of AI brings to media, society and politics, AI4Media aspires to become a centre of excellence and a wide-spread network of researchers across Europe and beyond. The focus is on delivering the next generation of core AI advances to serve the key sector of media, to ensure that European ethical values are embedded in future AI deployments, and to reimagine AI as a crucial, beneficial enabling technology in the service of society and media.

The AI4Media consortium comprises 30 leading partners in the areas of AI and media (9 universities, 9 research centres, 12 industrial partners) and 35 associate members. The aim is to establish a networking infrastructure to bring together the currently fragmented European AI landscape in the

field of media, and to foster deeper and long-running interactions between academia and industry, including Digital Innovation Hubs. A research agenda will be defined for AI research in the media domain, and research and innovation will be conducted with respect to cutting-edge AI technologies within specific fields of media-related AI. AI4Media will also provide a targeted funding framework through open calls, to speed up the uptake of innovations developed within the network.

A PhD programme will further enhance links to industry and the fostering and exchange of talent, while helping to prevent brain drain outside Europe. The network will develop a set of use cases to demonstrate the impact of advances achieved in the media sector. The Centre of Excellence established by AI4Media, together with the surrounding

ecosystem, will provide a long-term basis for the support of AI excellence in Europe, with the aim of ensuring that ethical AI, guided by European values, assumes a global leading role in the field of Media.

Contact: Fabrizio Sebastiani, AIMH Lab
fabrizio.sebastiani@isti.cnr.it
https://ai4media.eu/

NAVIGATOR

An imaging biobank to precisely prevent and predict cancer, and facilitate the participation of oncologic patients to diagnosis and treatment
Co-funded by the Tuscany Region under the Par Fas Salute Toscana 2014-2020 program

It is believed that oncology can benefit immensely from a paradigm shift towards personalised medical solutions that account for the great heterogeneity and intra-variability of tumour manifestation and response to treatment. Quantitative imaging and imaging-guided interventions play a key role in this perspective, as they provide multi-parametric morphologic and functional information of value for personalised predictions and prognoses, and new insights into the mechanisms underlying a patient's response to therapy.

NAVIGATOR aims to boost precision medicine in oncology by advancing translational research based on quantitative imaging and multi-omics analyses, towards a better understanding of cancer biology, cancer care, and, more generally, cancer risk.

The project will deliver a technological solution relying on:

- an **open image Biobank**, collecting and preserving large amounts of quality, standardised image data and related omics data in a secure and privacy-preserving model. Information will include CT, MRI and PET data for various neoplasms, clinical data from regional healthcare services (i.e., from Regional Health Agencies), molecular and liquid biopsy data
- an **open-science oriented, virtual research environment**, available to medical researchers and general clinical stakeholders, processing the multi-omics data to (i) extract gold-standard and novel imaging biomarkers based on *Radiomics analyses*; and to (ii) create and test digital patient models, through

data analytics techniques, based on cancer phenotypes, stratified risks and responsiveness to therapy.

Three highly-impacting, solid neoplasms will be initially considered as use cases to populate the Biobank (>=1500 cases) and to advance clinical findings in their respect. However, the Biobank data model will be highly flexible to ensure its scalability to integrate other tumour types.

NAVIGATOR relies on a strong regional network of Hospitals and Research Institutions in Pisa, Florence and Siena, partnered with European universities (i.e., Cambridge and Bournemouth) to establish an international basis for the studies. ISTI-CNR plays a key role in the project, as the three Labs involved (i.e., NeMIS, HPC and SI) will lead the design and deployment of the Virtual Research Environment as well as of the AI algorithms for the *Radiomics analyses.*
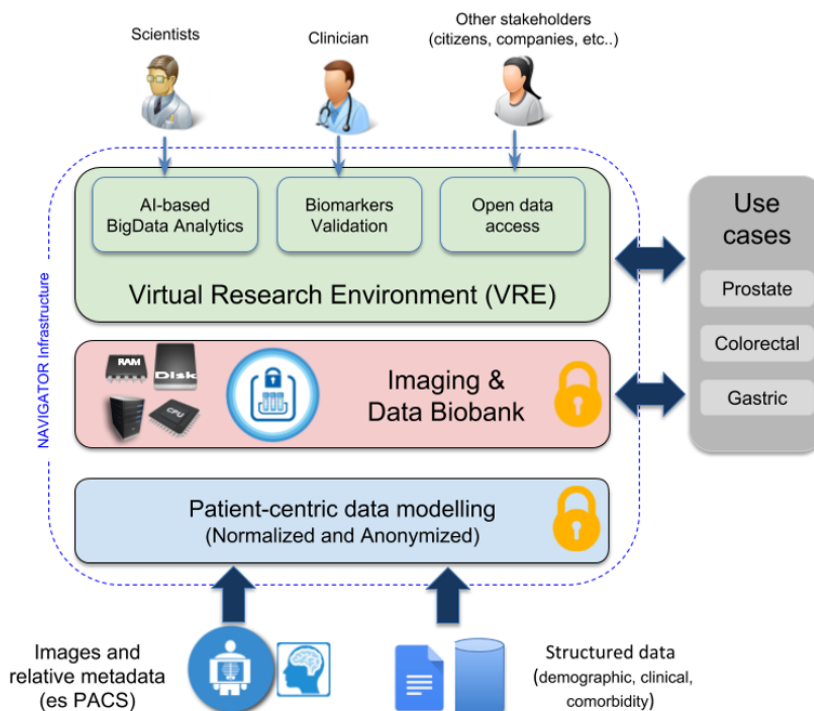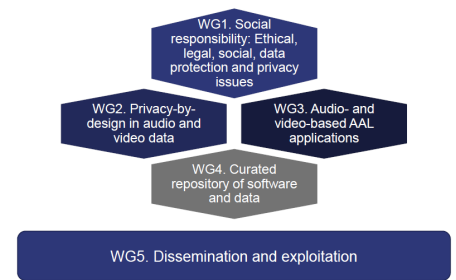

Contact: Sara Colantonio, SI Lab
Paolo Manghi, Paquale Pagano, NeMIS Lab
Emanuele Carlini, HPC Lab
sara.colantonio@isti.cnr.it
paolo.manghi@isti.cnr.it
pasquale.pagano@isti.cnr.it
emanuele.carlini@isti.cnr.it

GoodBrother

Network on privacy-aware audio- and video-based applications for active and assisted living - COST Program (COST Action CA19121)

Europe faces crucial challenges regarding health and social care due to demographic changes and the current economic context. Active and Assisted Living (AAL) is a possible solution. AAL aims at improving the health, quality of life, and wellbeing of older, impaired and frail people. AAL systems use different sensors to monitor the environment and society. Cameras and microphones are frequently used for AAL. They make it possible to monitor an environment and gather information, and are the most straightforward and natural way to describe events, persons, objects, actions, and interactions. Recent advances have given these devices the ability to 'see' and 'hear'. However, their use can be seen as intrusive by some end users (assisted persons, and professional and informal caregivers.) The General Data Protection Regulation (GDPR) establishes the obligation for technologies to meet the principles of data protection; data protection must be provided by default. AAL solutions must thus consider privacy-by-design methodologies in order to protect the fundamental rights of those being monitored. he aim of GoodBrother is to increase awareness of the ethical, legal, and privacy issues associated with audio- and video-based monitoring and to propose privacy-aware working solutions for assisted living, by creating an interdisciplinary community of researchers and industrial partners from different fields (computing, engineering, healthcare, law, sociology) and other stakeholders (users, policy makers, public services), stimulating new research and innovation. GoodBrother will offset the "Big Brother" sense of continuous monitoring by increasing user acceptance, exploiting these new solutions, and improving mar-


ket reach. The Signals & Images Lab of ISTI is involved in this project, contributing on privacy-preserving image and video analysis methods. Sara Colantonio is a member of the Management Committee of the Action and Co-leader of Working Group 3 - Audio- and Video-based AAL applications.

Contact: Sara Colantonio, SI Lab
sara.colantonio@isti.cnr.it

OSIRIS Follow on

The follow-on of the ESA GSTP Project OSIRIS "Optical/SAR data and System Integration for Rush Identification of Ship models"
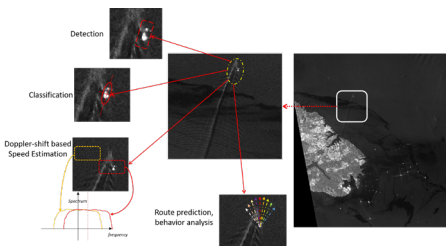
The main objective of OSIRIS-FO is to develop, implement, and integrate tools for Earth Observation Maritime Surveillance on an Open-Source platform treating both optical and Synthetic Aperture Radar (SAR) satellite imagery and Automatic Identification System (AIS) data from collaborative vessels.

This follow-on project aims to improve the functionalities of the OSIRIS modules for SAR image processing to raise their Tech-

nology Readiness Level (TRL) from TRL 6 "Technology demonstrated in relevant environment" to TRL8 "System complete and qualified" in order to support Copernicus

mission exploitation by the European Space Agency (ESA).

The partnership is composed of:

- Mapsat Srl, project leader, an Italian space company, developing the system for vessel detection;
- Sistemi Territoriali Srl, a CNR spinoff IT-consulting company, developing the WebGIS and the control panel modules;
- CNR, with the IIT and ISTI institutes, developing ship classification and kine-

matic estimation (ISTI), and ship route prediction and behaviour analysis (IIT).

The Signals and Images Lab contribution consists of improvements in the Ship Classification (SC) module for achieving accurate characterization from moderate-resolution SAR data, and in the Ship Kinematic Estimation (SKE) module for reliable vessel velocity estimation by Fourier analysis of Single-Look Complex (SLC) SAR images. All the studies and developments will be carried out taking the ESA's Sentinel-1 image features as reference.

Contact: Emanuele Salerno, SI Lab
emanuele.salerno@isti.cnr.it

Beyond Euclidean distance for error measurement in pedestrian indoor location sensing scenarios

G.M. Mendoza-Silva, J. Torres-Sospedra, F. Potortì, A. Moreira, S. Knauth, R. Berkvens, J. Huerta

Indoor Positioning Systems suffer from a lack of standard evaluation procedures enabling credible comparisons: this is one of the main challenges hindering their widespread market adoption. Traditionally, accuracy evaluation is based on positioning errors defined as the Euclidean distance between the true positions and the estimated positions. While Euclidean is simple, it ignores obstacles and floor transitions. In this paper, we describe procedures that measure a positioning error defined as the length of the pedestrian path that connects the estimated position to the true position. The procedures apply path-finding on floor maps using Visibility Graphs or Navigational Meshes for vector maps, and Fast Marching for raster maps. Multi-floor
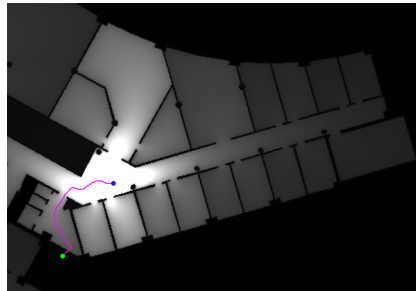
FM image output - FM examples. Blue and green dots represent origin and destination points, respectively. Magenta lines show the paths determined applying FM.

and multi-building paths use information on vertical in-building communication ways and outdoor paths. The proposed measurement procedures are applied to position estimates provided by the Indoor Positioning Systems that participated in the EvAAL-ETRI 2015 competition. Procedures are compared in terms of pedestrian path realism, indoor model complexity, path computation time and error magnitudes. The Visibility Graphs algorithm computes shortest distance paths; Navigational Meshes produces very similar paths with significantly shorter computation time; Fast Marching computes longer, more natural-looking paths at the expense of longer computation time and memory size. The 75th percentile of the measured error differs among the methods from 2.2 m to 3.7 m across the evaluation sets.

Discovering location based services: a unified approach for heterogeneous indoor localization systems

F. Furfari, A. Crivello, P. Baronti, P. Barsocchi, M. Girolami, F. Palumbo, D. Quezada-Gaiborb, G.M. Mendoza Silva, J. Torres-Sospedra

The technological solutions and communication capabilities offered by the Internet of Things paradigm, in terms of raising availability of wearable devices, the ubiquitous internet connection, and the presence on the market of service-oriented solutions, have allowed a wide proposal of Location Based Services (LBS). In a close future, we foresee that companies and service providers will have developed reliable solutions to address indoor positioning, as basis for use-ful location-based services. These solutions will be different from each other and they will adopt different hardware and processing techniques. This paper describes the proposal of a unified approach for Indoor Localization Systems that enables the cooperation between heterogeneous solutions and their functional modules. To this end, we designed an integrated architecture that, abstracting its main components, allows a seamless interaction among them. Finally, we present a working prototype of such architecture, which is based on the popular Telegram application for Android, as an integration demonstrator. The integration of the three main phases – namely the discovery phase, the User Agent self-configuration, and the indoor map retrieval/rendering – demonstrates the feasibility of the proposed integrated architecture.

Sensing social interactions through BLE beacons and commercial mobile devices

M. Girolami, F. Mavilia, Franca Delmastro
Pervasive and Mobile Computing, vol. 67. Elsevier, 2020.
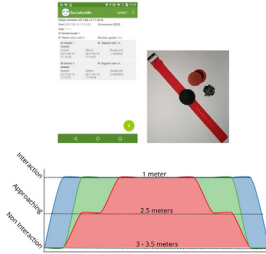
Hardware and software setting of the SocializeME project.

Wearable sensing devices can provide high-resolution data useful to characterise and identify complex human behaviours. Sensing human social interactions through wearable devices represents one of the emerging field in mobile social sensing, considering their impact on different user categories and on different social contexts. However, it is important to limit the collection and use of sensitive information characterising individual users and their social interactions in order to maintain the user compliance. For this reason, we decided to focus mainly on physical proximity and, specifically, on the analysis of BLE wireless signals commonly used by commercial mobile devices (see the Figure below). In this work, we present the SocializeME framework designed to collect proximity information and to detect social interactions through heterogeneous personal mobile devices. We also present the results of an experimental data collection campaign conducted with real users, highlighting technical limitations and performances in terms of quality of RSS, packet loss, and channel symmetry, and how they are influenced by different configurations of the user's body and the position of the personal device. Specifically, we obtained a dataset with more than 820.000 Bluetooth signals (BLE beacons) collected, with a total monitoring of over 11h. The dataset collected reproduces 4 different configurations by mixing two user posture's layouts (standing and sitting) and different positions of the receiver device (in hand, in the front pocket and in the back pocket). The large number of experiments in those different configurations, well cover the common way of holding a mobile device, and the layout of a dyad involved in a social interaction. We also present the results obtained by SME-D algorithm, designed to automatically detect social interactions based on the collected wireless signals, which obtained an overall accuracy of 81.56% and F-score 84.7%. The collected and labelled dataset is also released to the mobile social sensing community in order to evaluate and compare new algorithms.

Timed service contract automata

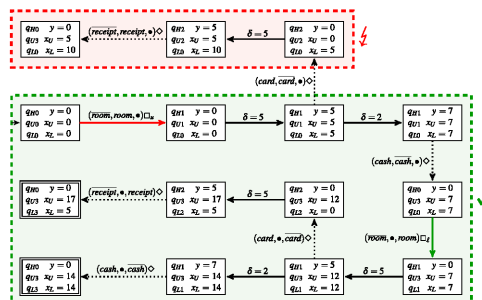D. Basile, M.H. ter Beek, A. Legay
Innovations in Systems and Software Engineering, vol. 16. Springer, 2020.

We equip a recently developed model for the specification of service contracts with real-time constraints. Service contracts offer a means to define the behavioural compliance of a composition of services, typically dictated in a service-level agreement (SLA), as the fulfilment of all service requests through service offers. Depending on their granularity, SLAs vary according to the level of criticality of the involved services and also contain real-time aspects, like the services' response or expiration time. A standard method to refine a spurious service composition into a compliant one is via the synthesis of a safe orchestration, in the form of the most permissive controller from supervisory control theory. Ideally, safe orchestrations solve competition among matching service requests and offers, in light of their criticali-

Excerpt of the behaviour of the composition of three timed service contract automata: the green fragment is allowed in the safe orchestration, while the red fragment is not.

ties and their timing constraints, in the best possible way. In this paper, we introduce timed service contract automata as a novel formal model for service contracts with real-time constraints on top of services with varying levels of criticality. We also define a means to efficiently compute their composition and their safe orchestration, using the concept of zones from timed games. The innovations of our contribution are illustrated by intuitive examples and by a preliminary evaluation.

Synthesis of orchestrations and choreographies: bridging the gap between supervisory control and coordination of services

D. Basile, M.H. ter Beek, R. Pugliese
Logical Methods in Computer Science, vol. 16. LMCS, 2020.

We present a number of contributions to bridging the gap between supervisory control theory and coordination of services in order to explore the frontiers between coordination and control systems. Firstly, we modify the classical synthesis algorithm from supervisory control theory for obtaining the so-called most permissive controller in order to synthesise orchestrations and choreographies of service contracts formalised as contract automata. The key ingredient to make this possible is a novel notion of controllability. Then, we present an abstract parametric synthesis algorithm and show that it generalises the classical synthesis as well as the orchestration and choreography syntheses. Finally, through the novel abstract synthesis, we show that the concrete syntheses are in a refinement order. A running example from the service domain illustrates our contributions.
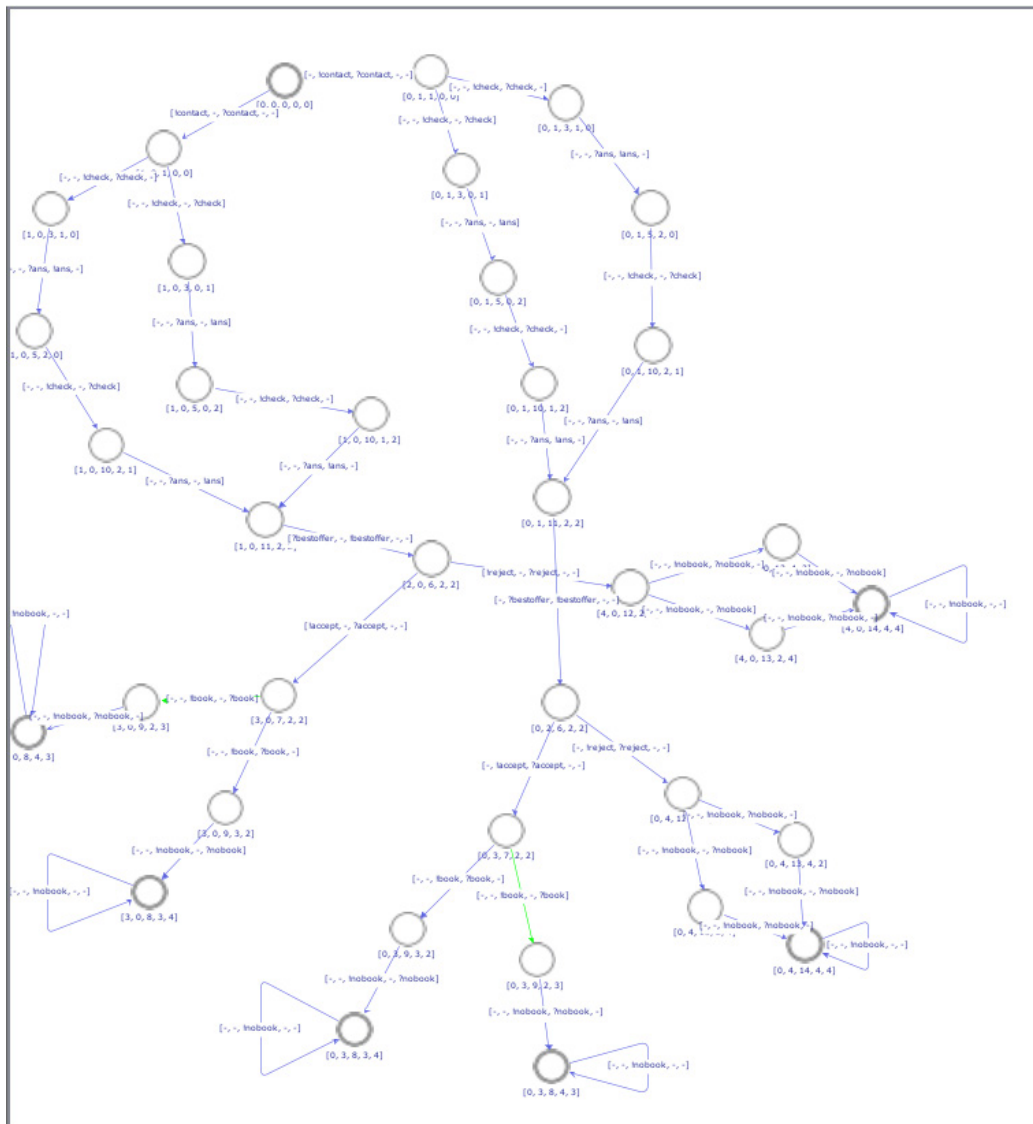
A snapshot of the accompanying tool FMCAT at work (cf. https://youtu.be/W0BHIgQEhIk).

The 2020 expert survey on formal methods

H. Garavel, M.H. ter Beek, J. van de Pol
Lecture Notes in Computer Science, vol. 12327. Springer, 2020.

This lengthy report has been organised to celebrate the 25th anniversary of the International Conference on Formal Methods for Industrial Critical Systems (FMICS), which was founded by researchers from the CNR Institute of Information Science and Technologies "A. Faedo" (ISTI). The survey addresses 30 questions on the past, present, and future of formal methods in research, industry, and education. A stunning 130 high-profile experts in formal methods (among whom three

Turing award winners, all three recipients of an FME Fellowship Award and thirteen CAV Award winners, as well as many recipients of other prizes and distinctions in computer science) participated in this survey. The report analyses their answers and comments, and moreover presents a collection of 111 position statements provided by these experts. The survey is both an unprecedented effort to gather the collective knowledge of the formal methods community, and a fam-

ily picture of key players in the field of formal methods. Formal methods are mathematics-based techniques for the specification, development, and (manual or automated) verification of software and hardware systems. The Formal Methods and Tools (FMT) laboratory of ISTI is an internationally renowned research group in this field, as witnessed by its involvement in this study.

An automated framework for continuous development and testing of access control systems

S. Daoudagh, F. Lonetti, E. Marchetti
Journal of Software: Evolution and Process, in press. Wiley, 2020.

Automated testing in DevOps represents a key factor for providing fast release of new software features assuring quality delivery. In this work, we introduce DOXAT (DevOps XAcml Testing), an automated framework for continuous development and testing of access control mechanisms based on the XACML standard. It leverages mutation analysis for the selection and assessment of the test strategies and provides automated facilities for test oracle definition, test execution, and results analysis, in order to speed up and automate the Plan, Code, Build, and Test phases of DevOps process.

The achievement and realization of the main goal of this work required different activities that span from the integration of existing tools, the implementation of new components, and the final evaluation.

In particular, the original contributions of the paper are:

- the definition and the implementation of DOXAT framework;
- the integration of existing tools in DOXAT for the test case generation, for the PDP code mutants derivation, and for

XACML-based oracle definition;

- an overview of the usage of DOXAT for the continuous development and testing of the PDP during the DevOps process; and
- a first evaluation to show the effectiveness of DOXAT in development and testing of an example PDP.
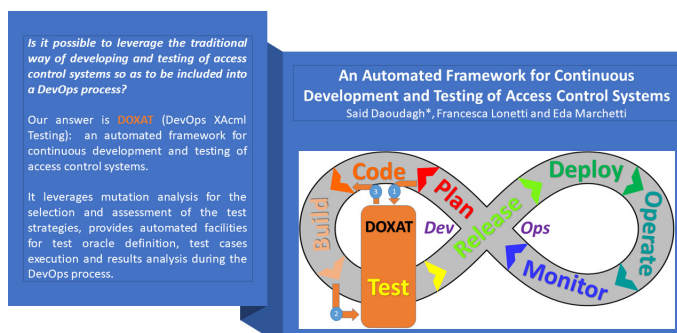
DOXAT inside the DevOps process.

Learning-to-rank vs ranking-to-learn: strategies for regression testing in continuous integration

A. Bertolino, A. Guerriero, B. Miranda, R. Pietrantuono, S. Russo
ICSE '20 - 42nd International Conference on Software Engineering. ACM, 2020.

Continuous Integration (CI) practices in large industrial settings pose specific requirements on test selection and prioritization for regression testing, due to the frequent commits, with short inter-commit times and few changes with respect to the codebase size. To address such requirements, some Machine Learning (ML) techniques have been proposed, as an alternative to deterministic approaches. Two broad strategies for ML-based prioritization are learning-to-rank and what we call ranking-to-learn (i.e., reinforcement learning). Various ML algorithms can be applied in each strategy. In this paper we introduce ten of such algorithms for adoption in CI practices, and perform a comprehensive study comparing them against each other using subjects from the Apache Commons project. We analyze the influence of several features of the code under test and of the test process. The results allow to draw criteria to support testers in selecting and tuning the technique that best fits their context. We find that a valuable approach under short inter-commit times is to run a ML prioritization algorithm after selection, so that it acts on a small problem size and on test cases relevant for that commit.

In the plethora of ML algorithms to choose from, our suggestion is to first decide the learning strategy and, possibly, the category (ensemble or non-esemble), and then opt for the specific algorithm. The choice depends on the requirements for prioritization, in terms of desired ranking effectiveness and efficiency, tolerable sensitivity to code features, and on the CI process features.
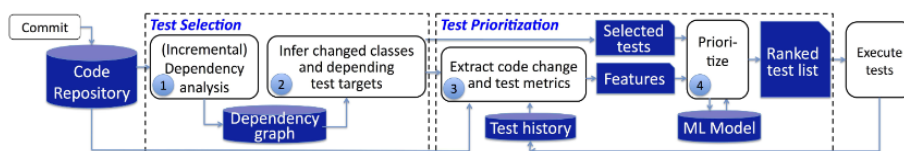
Overview of the test selection and prioritization process.

Requirements management in automotive: an empirical study on process improvement areas

F. Falcini, G. Lami
IEEE 28th International Requirements Engineering Conference. IEEE, 2020.

In the automotive domain, the development of software-intensive components is mainly demanded to specialized suppliers. Many car manufacturers (a.k.a. OEMs - Original Equipment Manufacturer) request suppliers to measure and, eventually, improve their development process by applying process assessment models such as Automotive SPICE. Automotive SPICE is the reference model for software development process evaluation and improvement in automotive. Automotive SPICE provides a set of requested software development life-cycle practices, including requirements management. Requirements management is a crucial issue in automotive because of the high volatility of requirements during projects and the need for interactions among different stakeholders. This empirical study aims at contributing to the identification of what are the most frequent weaknesses in requirements management in automotive. In the capacity of Automotive SPICE assessors, the authors have evaluated requirement management practices based on the evidence gathered from real industrial development projects during a significant number of assessments performed at several organizations worldwide. This paper intends to derive a picture of the state-of-the-practice of requirements management in automotive, focusing on the development of software-intensive automotive components. The purpose is to provide researchers and practitioners with a reference for improvement initiatives to solve requirement engineering problems. This study's principal originality is the use of real data from real software development projects in automotive. In literature, the empirical studies addressing similar topics are mainly based on literature reviews and surveys made through questionnaires.

Cross-resolution learning for face recognition

F.V. Massoli, G. Amato, F. Falchi
Image and Vision Computing, vol. 99. Elsevier 2020.

Convolutional Neural Network models have reached extremely high performance on the Face Recognition task. Mostly used datasets, such as VGGFace2, focus on gender, pose, and age variations, in the attempt of balancing them to empower models to better generalize to unseen data. Nevertheless, image resolution variability is not usually discussed, which may lead to a resizing of 256 pixels. While specific datasets for very low-resolution faces have been proposed, less attention has been paid on the task of cross-resolution matching. Hence, the discrimination power of a neural network might seriously degrade in such a scenario. Surveillance systems and forensic applications are particularly susceptible to this problem since, in these cases, it is common that a low-resolution query has to be matched against higher-resolution galleries. Although it is always possible to either increase the resolution of the query image or to reduce the size of the gallery (less frequently), to the best of our knowledge, extensive experimentation of cross-resolution matching was missing in the recent deep learning-based literature. In the context of low- and cross-resolution Face Recognition, the contribution of our work is fourfold: i) we proposed a training procedure to fine-tune a state-of-the-art model to empower it to extract resolution-robust deep features; ii) we conducted an extensive test campaign by using high-resolution datasets (IJB-B and IJB-C) and surveillance-camera-quality datasets (QMUL-SurvFace, TinyFace, and SCface) showing the effectiveness of our algorithm to train a resolution-robust model; iii) even though our main focus was the cross-resolution Face Recognition, by using our training algorithm we also improved upon state-of-the-art model performances considering low-resolution matches; iv) we showed that our approach could be more effective concerning preprocessing faces with super-resolution techniques.
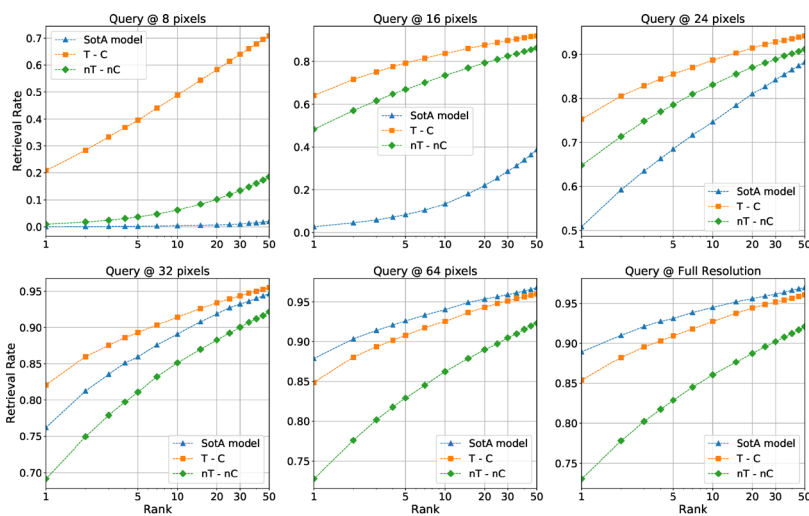
CMC curves. The higher the curve the better the discrimination power of the model. Each plot corresponds to a different query resolution.

2D t-SNE embeddings for 20 different identities randomly extracted from the VGGFace2 dataset. All the images were down-sampled to a resolution of 8 pixels. Left: "Base Model". Right: "T-C" model.
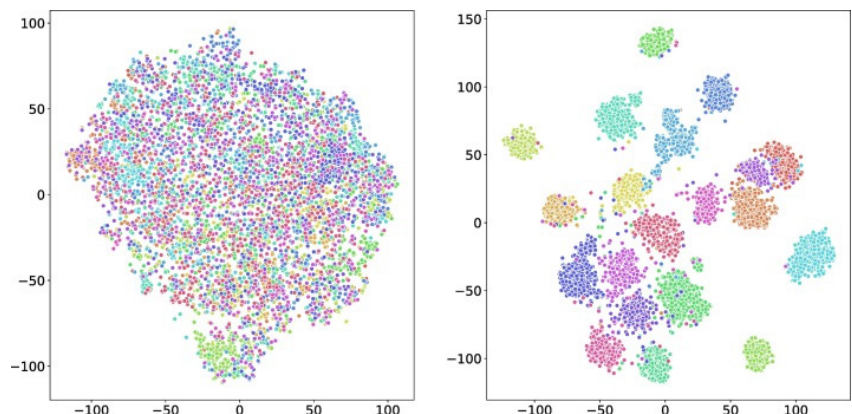
Detection of face recognition adversarial attacks

F.V. Massoli, F. Carrara, G. Amato, F. Falchi
Computer Vision and Image Understanding, in press. Elsevier, 2020.

Deep Learning methods have become state-of-the-art for solving tasks such as Face Recognition (FR). Unfortunately, despite their success, it has been pointed out that these learning models are exposed to adversarialinputs – images to which an imperceptible amount of noise for humans is added to maliciously fool a neural network – thus limiting their adoption in sensitive real-world applications. While it is true that an enormous effort has been spent to train robust models against this type of threat, adversarial detection techniques have recently started to draw attention within the scientific community. The advantage of using a detection approach is that it does not require to re-train any model; thus, it can be added to any system. In this context, we present our work on adversarial detection in forensics mainly focused on detecting attacks against FR systems in which the learning model is typically used only as features extractor. Thus, training a more robust classifier might not be enough to counteract the adversarial threats. In this frame, the contribution of our work is four-fold: (i) we test our proposed adversarial detection approach against classification attacks, i.e., adversarial samples crafted to fool an FR neural network acting as a classifier; (ii) using a k-Nearest Neighbor (k-NN) algorithm as a guide, we generate deep features attacks against an FR system based on a neural network acting as features extractor, followed by a similarity-based procedure which returns the query identity; (iii) we use the deep features attacks to fool an FR system on the 1:1 face verification task, and we show their superior effectiveness with respect to classification attacks in evading such type of system; (iv) we use the detectors trained on the classification attacks to detect the deep features attacks, thus showing that such approach is generalizable to different classes of offensives.
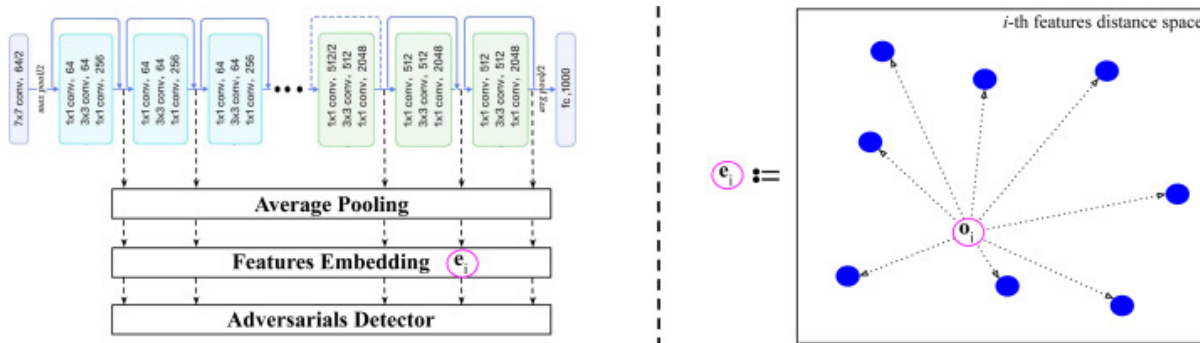
Schematic view of the proposed detection algorithm. Left: model architecture and embedding procedure. The magenta circle represents a single embedding vector . Right: representation of the th features distance space. Given a features map , we evaluate its embedding as the vector whose components are the distances between the current features and the pivots (blue circles), centroids or medoids, of each class. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Adversarial examples for two different values of the threshold applied while solving the optimization problem. Top: δ= 5.0. Bottom: δ= 10.0.

Evaluation measures for quantification: an axiomatic approach

F.Sebastiani
Information Retrieval Journal, vol. 23. Springer, 2020.

Quantification is the task of estimating, given a set σ of unlabelled items and a set of classes $C=\{c_1,...,c_{|C|}\}$, the prevalence (or "relative frequency") in σ of each class $c_i \in C$. While quantification may in principle be solved by classifying each item in σ and counting how many such items have been labelled with c_i, it has long been shown that this "classify and count" method yields suboptimal quantification accuracy. As a result, quantification is no longer considered a mere byproduct of classification, and has evolved as a task of its own. While the scientific community has devoted a lot of attention to devising more accurate quantification methods, it has not devoted much to discussing what properties an evaluation measure for quantification (EMQ) should enjoy, and which EMQs should be adopted as a result. This paper lays down a number of interesting properties that an EMQ may or may not enjoy, discusses if (and when) each of these properties is desirable, surveys the EMQs that have been used so far, and discusses whether they enjoy or not the above properties. As a result of this investigation, some of the EMQs that have been used in the literature turn out to be severely unfit, while others emerge as closer to what the quantification community actually needs. However, a significant result is that no existing EMQ satisfies all the properties identified as desirable, thus indicating that more research is needed in order to identify (or synthesize) a truly adequate EMQ.
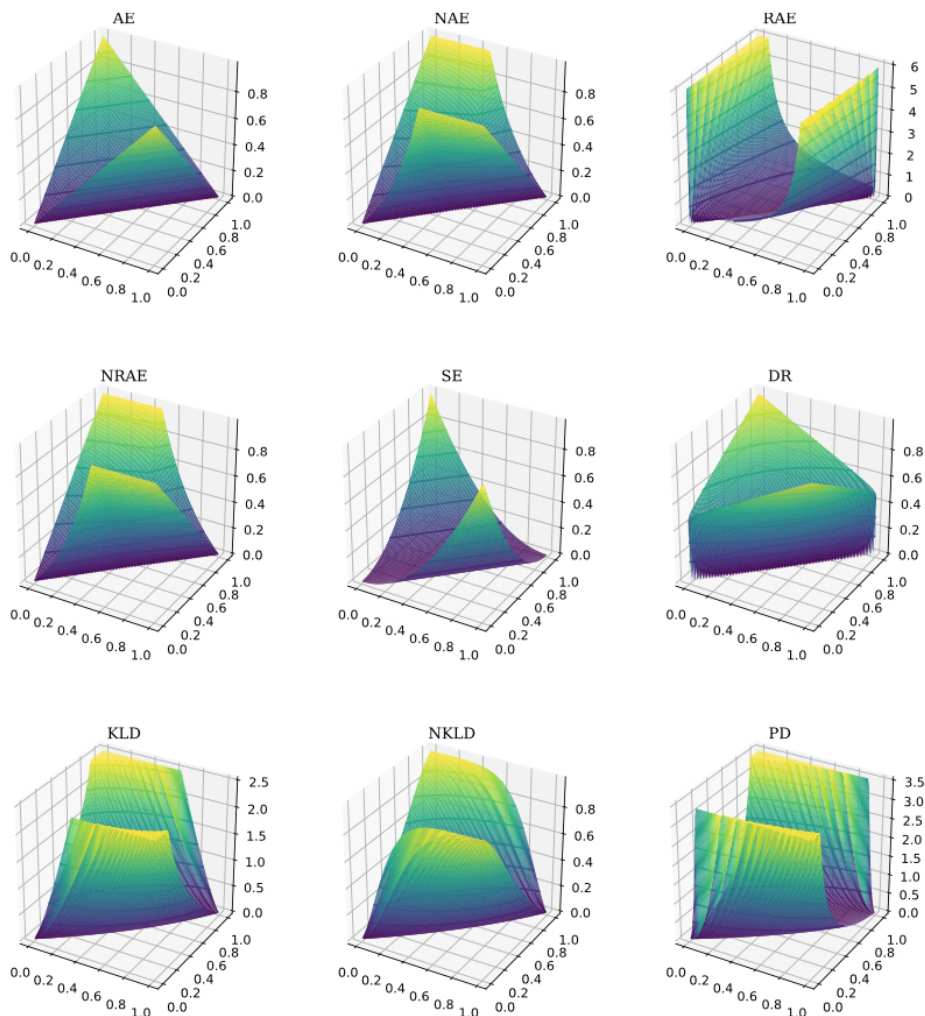
3D plots (for a binary quantification task) for the nine EMQs studied in this paper; $p(c_1)$ and $p(c_2)$ are represented as x and $(1-x)$, respectively, while $p\hat{}(c_1)$ and $p\hat{}(c_2)$ are represented as y and $(1-y)$; error is represented as z (higher values of z represent higher error).

A critical reassessment of the Saerens-Latinne-Decaestecker algorithm for posterior probability adjustment

A. Esuli, A. Molinari, F. Sebastiani

We critically re-examine the Saerens-Latinne-Decaestecker (SLD) algorithm, a well-known method for es- timating class prior probabilities ("priors") and adjusting posterior probabilities ("posteriors") in scenarios characterized by distribution shift, i.e., difference in the distribution of the priors between the training and the unlabelled documents. Given a machine learned classifier and a set of unlabelled documents for which the classifier has returned posterior probabilities and estimates of the prior probabilities, SLD updates them both in an iterative, mutually recursive way, with the goal of making both more accurate; this is of key importance in downstream tasks such as single-label multiclass classification and cost-sensitive text classification. Since its publication, SLD has become the standard algorithm for improving the quality of the posteriors in the pres- ence of distribution shift, and SLD is still considered a top con- tender when we need to estimate the priors (a task that has become known as "quantifi- cation"). However, its real effectiveness in improving the quality of the posteriors has been questioned. We here present the re- sults of systematic experiments conducted on a large, publicly available dataset, across multiple amounts of distribution shift and multiple learners. Our experiments show that SLD improves the quality of the poste- rior probabilities and of the estimates of the prior probabilities, but only when the num- ber of classes in the classification scheme is very small and the classifier is calibrated. As the number of classes grows, or as we use non-calibrated classifiers, SLD converges more slowly (and often does not converge at all), performance degrades rapidly, and the impact of SLD on the quality of the prior es- timates and of the posteriors becomes nega- tive rather than positive.

Representing narratives in digital libraries: the narrative ontology

C. Meghini, V. Bartalesi, D. Metilli

Digital Libraries (DLs), especially in the Cultural Heritage domain, are rich in narra- tives. Every digital object in a DL tells some kind of story, regardless of the medium, the genre, or the type of the object. However, DLs do not offer services about narratives, for example it is not possible to discover a narrative, to create one, or to compare two narratives. Certainly, DLs offer discovery functionalities over their contents, but these services merely address the objects that car- ry the narratives (e.g. books, images, audio- visual objects), without regard for the narra- tives themselves. The present work aims at introducing narratives as first-class citizens
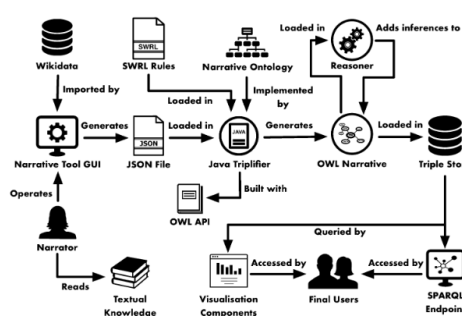
A view of the architecture of NBVT showing also the interactions between its components and its users.

in DLs, by providing a formal expression of what a narrative is. In particular, this paper presents a conceptualization of the domain of narratives, and its specification through the Narrative Ontology (NOnt for short), ex- pressed in first-order logic. NOnt has been implemented as an extension of three stan- dard vocabularies, i.e. the CIDOC CRM, FR- BRoo, and OWL Time, and using the SWRL rule language to express the axioms. An ini- tial validation of NOnt has been performed in the context of the Mingei European proj- ect, in which the ontology has been applied to the representation of knowledge about Craft Heritage.

The impact of serious games with humanoid robots on mild cognitive impairment older adults

M. Manca, F. Paternò, C. Santoro, E. Zedda, C. Braschi, R. Franco, A. Sale
International Journal of Human-Computer Studies, in press. Elsevier, 2020.

The number of Mild Cognitive Impairment (MCI) older adults is increasing; thus, it becomes more and more important to provide them with support to avoid, or at least slow down, their cognitive decline. To this end, interactive serious games can play an important role. So far, most of them have been deployed through tablets, which represent a cost-effective solution, yet offering only limited possibilities for truly engaging such users in a multimodal manner. However, emerging humanoid robots, through their physical embodiment and human-like attributes, including facial expressions and body language, may open up new possibilities in more effectively engaging MCI older adults during repetitive cognitive training. We present a study aiming to better understand the impact of humanoid robots in supporting serious games for such users. In particular, we investigate how seniors with Mild Cognitive Impairments relate to and perceive serious games accessed through humanoid robots, as part of a training programme aimed to improve their cognitive status. For this purpose, two versions of a music-based memory game have been designed by a multi-disciplinary team, one for humanoid robots and one for tablets. We report on its use during a between-subject study that involved MCI seniors, and discuss their experience. The results show that the robot was received with more enthusiasm by the older adults, thus improving their level of engagement.
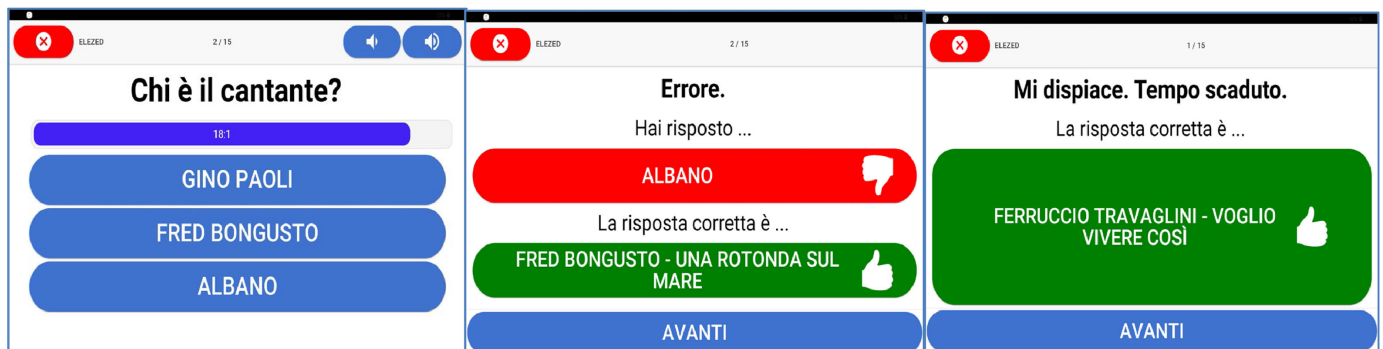
The application presenting a question (a), providing feedback for an incorrect answer (b), and for an occurred timeout (c).
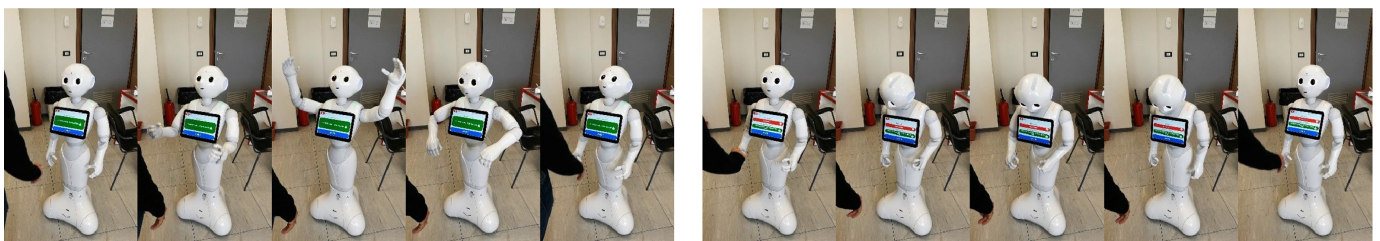
Example of positive (left part) and negative feedback (right part) from the robot.

Realizing virtual research environments for the agri-food community: the AGINFRA PLUS experience

M. Assante, A. Boizet, L. Candela,D. Castelli, R. Cirillo, G. Coro G., E. Fernández, M. Filter, L. Frosini, T. Georgiev, G. Kakaletris, P. Katsivelis, R. Knapen, L. Lelii, R.M. Lokers, F. Mangiacrapa, N. Manouselis, P. Pagano, G. Panichi, L. Penev, F. Sinibaldi

The enhancements in IT solutions and the open science movement are injecting changes in the practices dealing with data collection, collation, processing, analytics, and publishing in all the domains, including agri-food. However, in implementing these changes one of the major issues faced by the agri-food researchers is the fragmentation of the "assets" to be exploited when per-forming research tasks, for example, data of interest are heterogeneous and scattered across several repositories, the tools model-ers rely on are diverse and often make use of limited computing capacity, the publish-ing practices are various and rarely aim at making available the "whole story" including datasets, processes, and results. This paper presents the AGINFRA PLUS endeavor to overcome these limitations by providing re-searchers in three designated communities with Virtual Research Environments facili-tating the use of the "assets" of interest and promote collaboration.
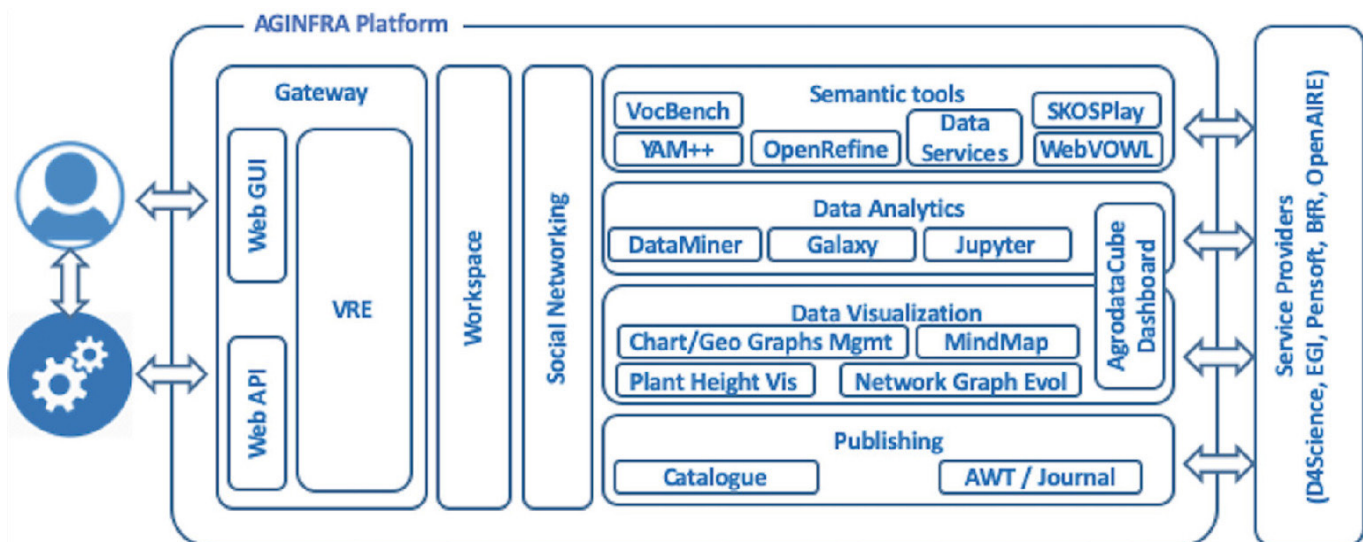
AGINFRA PLUS VRE Platform Architecture.

Ranking places in attributed temporal urban mobility networks

M. Nanni, L. Tortosa, J.F.Vicent, G. Yeghikyan
Plos One, vol. 15. Public Library Science, 2020.

Drawing on the recent advances in complex network theory, urban mobility flow patterns, typically encoded as origin-destination (OD) matrices, can be represented as weighted directed graphs, with nodes denoting city locations and weighted edges the number of trips between them. Such a graph can further be augmented by node attributes denoting the various socio-economic

characteristics at a particular location in the city. In this paper, we study the spatio-temporal characteristics of "hotspots" of different types of socio-economic activities as characterized by recently developed attribute-augmented network centrality measures within the urbanODnetwork. The workflow of the proposed paper comprises the construction of temporalODnetworks

using two custom data sets on urban mobility in Rome and London, the addition of socio-economic activity attributes to theOD-network nodes, the computation of network centrality measures, the identification of "hotspots" and, finally, the visualization and analysis of measures of their spatio-temporal heterogeneity. Our results show structural similarities and distinctions between the spatial patterns of different types of human activity in the two cities. Our approach produces simple indicators thus opening up opportunities for practitioners to develop tools for real-time monitoring and visualization of interactions between mobility and economic activity in cities.

DOI: 10.1371/journal.pone.0239319
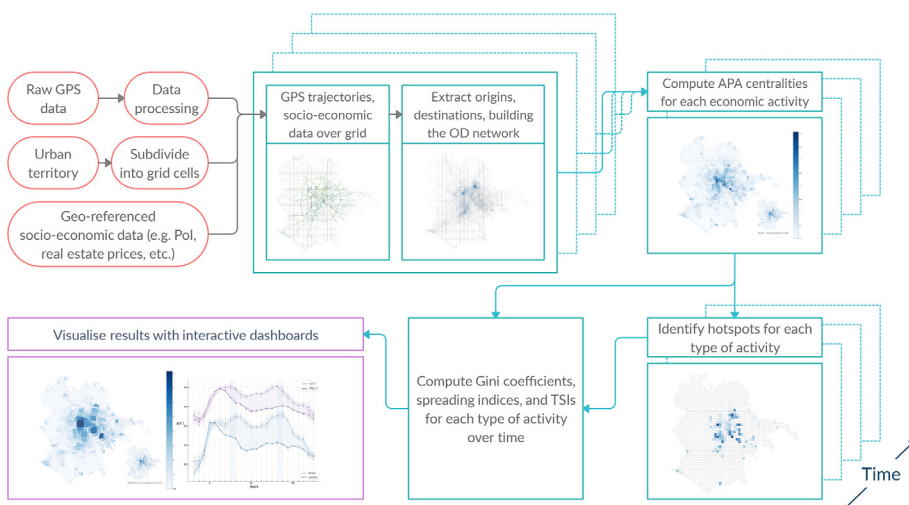
Flowchart of the methodology.

Modeling adversarial behavior against mobility data privacy

R. Pellungrini, L. Pappalardo, F. Simini, A. Monreale
IEEE Transactions on Intelligent Transportation Systems, in press. IEEE, 2020.

Privacy risk assessment is a crucial issue in any privacy-aware analysis process. Traditional frameworks for privacy risk assessment systematically generate the assumed knowledge for a potential adversary, evaluating the risk without realistically modelling the collection of the background knowledge used by the adversary when performing the attack. In this work, we propose Simulated

Privacy Annealing (SPA), a new adversarial behavior model for privacy risk assessment in mobility data. We model the behavior of an adversary as a mobility trajectory and introduce an optimization approach to find the most effective adversary trajectory in terms of privacy risk produced for the individuals represented in a mobility data set. We use simulated annealing to optimize the move-

ment of the adversary and simulate a possible attack on mobility data. We finally test the effectiveness of our approach on real human mobility data, showing that it can simulate the knowledge gathering process for an adversary in a more realistic way.

DOI: 10.1109/TITS.2020.3021911

On optimally partitioning variable-byte codes

G.E. Pibiri, R. Venturini
IEEE Transactions on Knowledge and Data Engineering, vol. 32. IEEE, 2020.

The ubiquitous Variable-Byte encoding is one of the fastest compressed representation for integer sequences. However, its compression ratio is usually not competitive with other more sophisticated encoders, especially when the integers to be compressed are small which is the typical case for inverted indexes. This paper shows that the compression ratio of Variable-Byte can be improved by 2X by adopting a partitioned representation of the inverted lists. This makes Variable-Byte surprisingly competitive in space with the best bit-aligned encoders, hence disproving the folklore belief that Variable-Byte is space-inefficient for inverted index compression.

Despite the significant space savings, we show that our optimization almost comes for free, given that: we introduce an optimal partitioning algorithm that does not affect indexing time because of its linear-time complexity; we show that the query processing speed of Variable-Byte is preserved, with an extensive experimental analysis and comparison with several other state-of-the-art encoders.
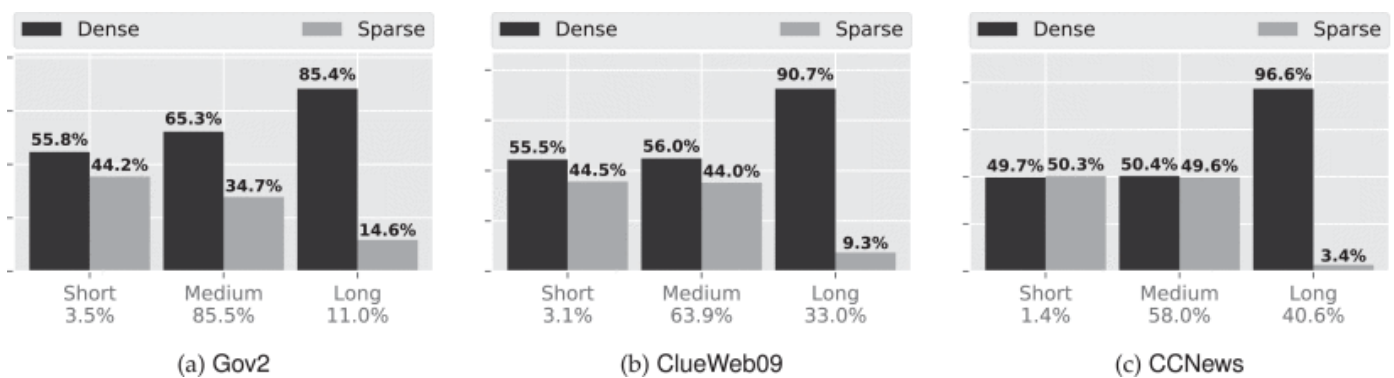
Percentage of integers belonging to dense and sparse regions of the inverted lists for the tested datasets. The inverted lists have been clustered by size into three categories: short (size < 10K); medium (10K ≤ size < 7M); long (size ≥ 7M). Below each category we also indicate the percentage of integers belonging to its inverted lists.

Practical trade offs for the prefix sum problem

G.E. Pibiri, R. Venturini
Software: Practice and Experience, in press. Wiley, 2020.

Given an integer array A, the prefix-sum problem is to answer sum(i) queries that return the sum of the elements in A[0..i], knowing that the integers in A can be changed. It is a classic problem in data structure design with a wide range of applications in computing from coding to databases. In this work, we propose and compare practical solutions to this problem, showing that new trade-offs between the performance of queries and updates can be achieved on modern hardware.
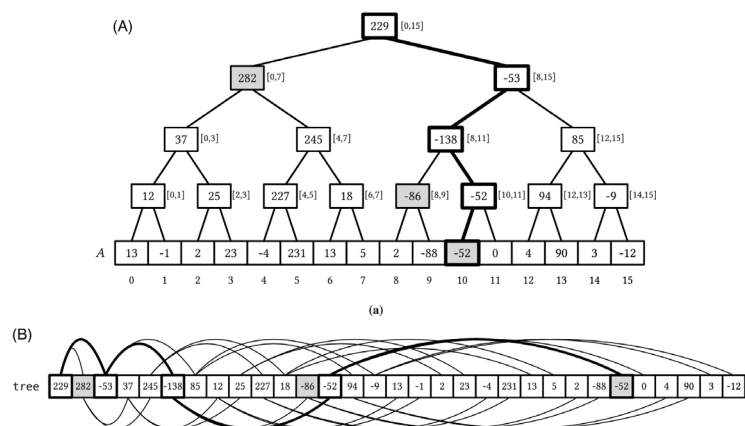
In (A), an example of Segment-Tree built from an input array A, with highlighted nodes belonging to the root-to-leaf path for answering sum(10). The shaded nodes are the ones accessed to compute the result, thus sum(10) = 282 − 86 − 52 = 144. In (B), the array-like representation of the same tree.

Scalable decentralized indexing and querying of multi-streams in the fog

P. Dazzi, M. Mordacchini
Journal of Grid Computing, vol. 18. OSA, 2020.

NOA-AID (Network Overlays for Adaptive information Aggregation, Indexing and Discovery on the fog) is an approach for decentralized indexing, aggregation and discovery of data belonging to streams. It is organized on two network layers. The upper layer is in charge of delivering an information discovery approach by providing a distributed index structure. The lower layer is devoted to resource aggregation based on epidemic protocols designed for highly dynamic environment, well suited to stream-oriented scenarios. It defines a flexible approach to express queries targeting highly heterogeneous data, as well as a self-organizing dynamic system allowing the optimal resolution of queries on the most suitable stream producers. The paper also presents a theoretical study and discusses the costs related to information management operations; it also gives an empirical validation of findings. Finally, it reports an extended experimental evaluation that demonstrated the ability of NOA-AID to be effective and efficient for retrieving information extracted from streams in highly-dynamic and distributed processing architectures.

NOA-AID architecture.
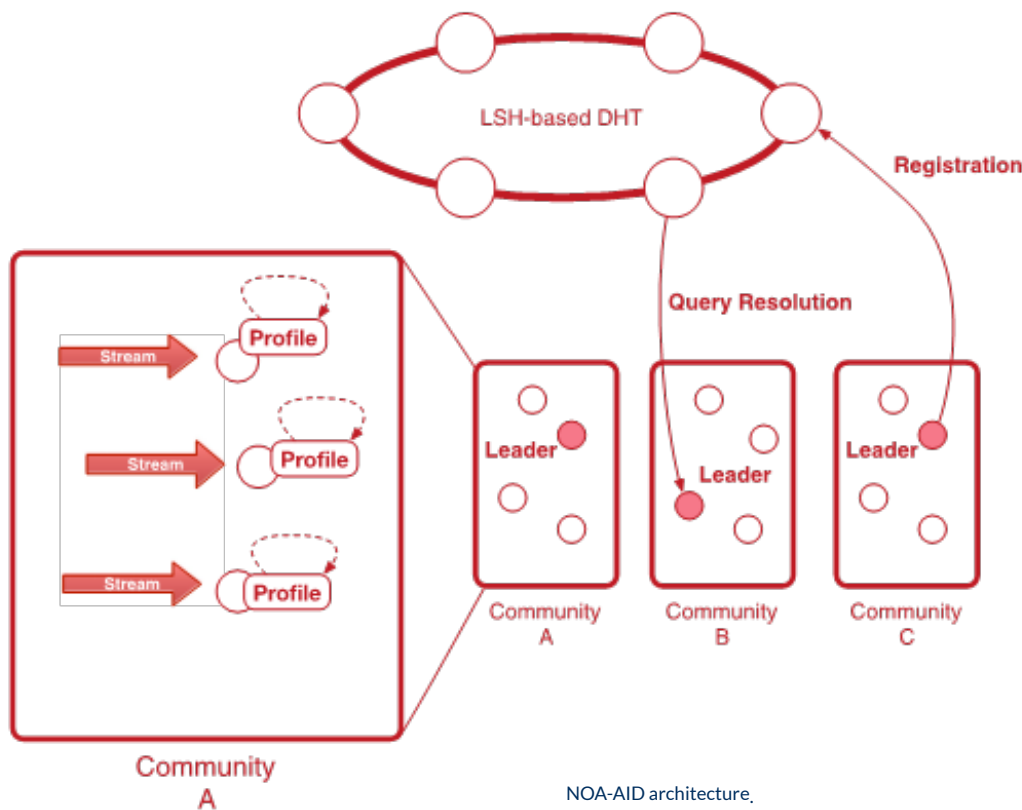
Weighting passages enhances accuracy

C.I. Muntean, F.M. Nardini, R. Perego, N. Tonellotto, O. Frieder
ACM Transactions on Information Systems, art. 11. ACM, 2020.

Believing that certain passages pose greater relevance to a given query and that the distribution of salient terms in the content of curated documents is collection dependent, we investigate how passage relevance can be exploited to improve effectiveness on domain-specific retrieval tasks, i.e., for news and web documents collections.

We observe that in curated documents the distribution of the occurrences of salient terms, e.g., terms with a high Inverse Document Frequency, is not uniform, and such terms are primarily concentrated towards the beginning and the end of the document.

Exploiting this observation, we propose a novel version of the classical BM25 weighting model, called BM25 Passage (BM25P), which scores query results by computing a linear combination of term statistics in the different portions of the document. We study a multiplicity of partitioning schemes
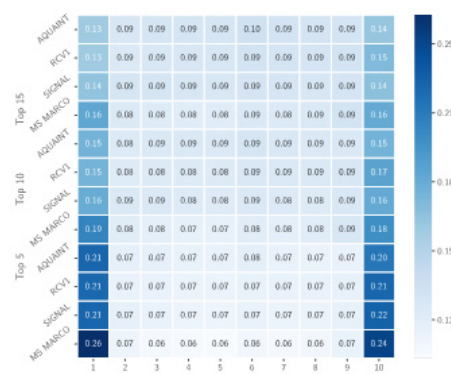
Probability distribution for the positions of salient terms occurrences in the documents for each of the four collections used.

of document content into passages and compute the collection-dependent weights associated with them on the basis of the distribution of occurrences of salient terms in documents. Moreover, we tune BM25P hyperparameters and investigate their impact on ad hoc document retrieval through fully reproducible experiments conducted using four publicly available datasets. Our findings demonstrate that our BM25P weighting model markedly and consistently outperforms BM25 in terms of effectiveness by up to 17.44% in NDCG@5 and 85% in NDCG@1, and up to 21% in MRR.

We perform a similar analysis to evaluate if our approach for passage weighting is general and effective also on different models. Consequently, we experiment with two different probabilistic weighting model families, namely, language models (LM) and models based on the divergence from randomness (DFR). For all cutoff values tested, all passage-enhanced weighting models clearly outperform the corresponding original weighting models.

High altitude mountain telemedicine

M. Martinelli, L. Bastiani, D. Moroni, S. Mrakic-Sposta, G. Giardini G., L. Pratali
Journal of Telemedicine and Telecare, in press. SAGE, 2020.

Introduction: An innovative teleconsultation platform has been designed, developed and validated between Summer 2017 and Winter 2018, in five mountain huts and in three remote outpatients clinical centres of the Italian region Valle d'Aosta of the Mont Blanc massif area.

Methods: An ad hoc video- conference system was developed within the framework of the e-Rés@mont (Interreg Alcotra) European project, to tackle general health problems and high altitude diseases (such as acute mountain sickness, high-altitude pulmonary and cerebral edema). The system allows contacting physicians at the main hospital in Aosta to perform a specific diagnosis and to give specific advice and therapy to the patients in an extreme environment out-hospital setting. At an altitude between 1500 and 3500 meters, five trained nurses performed clinical evaluations (anamnesis, blood pressure, heart rate, oxygen saturation), electrocardiographic and echography monitoring on both tourists and residents as necessary; all the collected data were sent to the physicians in Aosta.

Results: A total of 702 teleconsultation cases were performed: 333 dismissed (47%), 356 observed (51%), 13 immediate interventions (2%). In 30 cases the physicians decided there was no need for helicopter and ambulance rescue intervention and hospital admissions. The main physiological measures, the classified pathologies, the severe cases, and cost savings are described.

Discussion: The e-Rés@mont teleconsultation platform has been discussed in terms of treated cases, feasibility, proactivity in reducing complexities, direct and indirect advantages, and diagnostics help; moreover general and specific pros and cons have been debated, and future steps have been exposed.

Design guidelines for an interactive 3D model as a supporting tool for exploring a cultural site by visually impaired and sighted people

B. Leporini., V. Rossetti, F. Furfari, S. Pelagatti, A. Quarta
ACM Transactions on Accessible Computing, vol. 13. ACM, 2020.

Being able to explore and familiarise themselves with the structure and details of a cultural site before actually visiting it is fundamental for orienting visually impaired people during the visit; otherwise, it is particularly difficult to gain a global understanding of the structure and an overall impression of a square, a church, or a large monument. Our project addressed this problem by using low cost 3D models combined with audio descriptions to enable visually impaired users to explore the cultural site autonomously. Audio descriptions are organised into three groups (for historical, practical, and architectural information), and for each group, several tracks are recorded giving increasing levels of details. Users can easily navigate through the audio tracks to follow their tactile exploration by listening to the information they are most interested in. Relevant details are reproduced separately and linked to the main model via the audio tracks. A goal of our model is to enhance the understanding of the cultural site also for partially sighted as well as sighted people, making them able to appreciate the details of the architectural design using both visual and auditory senses. We exploited low-cost and partially open-source technologies, thus rendering our system easily replicable. We evaluated the interactive system with blind, partially sighted, and sighted users. Our user test confirmed the validity of our approach: (1) the 3D models and the tactile reproduction of details obtained via a low-cost 3D printing solution are well perceived by touch; (2) the semantic auditory information activated via perceptible buttons on demand and the different content levels for the audio tracks are suitable for an interactive, autonomous, and satisfying exploration; and (3) relevant details are well perceived. Finally, we propose guidelines to use in the 3D reproduction of buildings or large sites based on our experience.

Identifying value-increasing actions for cultural heritage assets through sensitivity analysis of multicriteria evaluation results
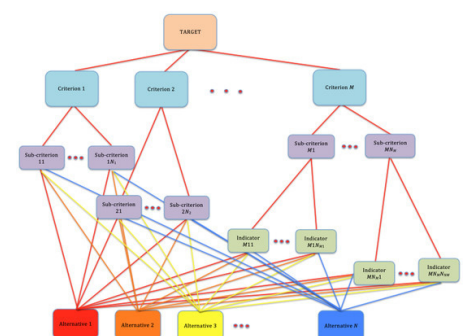
E. Salerno
Sustainability, vol. 12. MDPI, 2020.

This paper presents a brief overview of multicriteria decision making (MCDM) as applied to the evaluation of adaptive reuse projects for cultural heritage assets and proposes a strategy to plan interventions to increase their value. The value of an object can be defined from its fitness to fulfil specified objectives, its significance to the people who own or use it, its potential to produce revenues, and a host of other criteria depending on its nature. These criteria are often subjective, relying on judgements issued by several experts, stakeholders and decision makers.

This is why the MCDM methods need to formalize the problem so as to make it suitable to be treated quantitatively. Moreover, its sensitivity to variable opinions must be studied to check the stability of the result. We propose to leverage sensitivity analysis to identify the lines of intervention that promise to be the most effective to increase the value of the asset. A simulated example illustrates this strategy. This approach promises to be useful when assessing the sustainability of a reuse or redevelopment project in the cases where the final destination of the asset is still under examination.

A generic hierarchy for a problem structured by a hierarchical decision model (HDM). There are five hierarchical layers: the top layer (pink) includes the target decision; immediately below, there is the criterion layer (light blue), then the sub-criterion layer (purple) and the indicator layer (light green). At the bottom of the hierarchy, there are the alternative decisions, marked with different colors, which must be assigned relative scores with respect to all the elements of the lowest hierarchical layers.

Loopycuts: practical feature-preserving block decomposition for strongly hex-dominant meshing

Livesu M, Pietroni N., Puppo E., Sheffer A., Cignoni P.
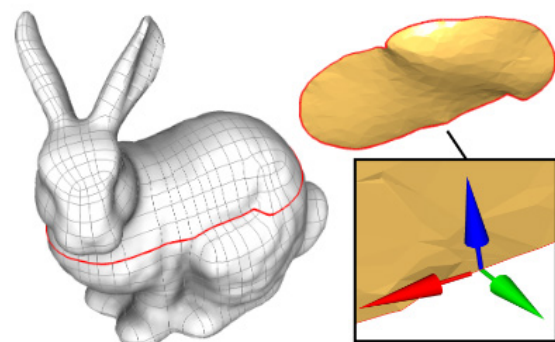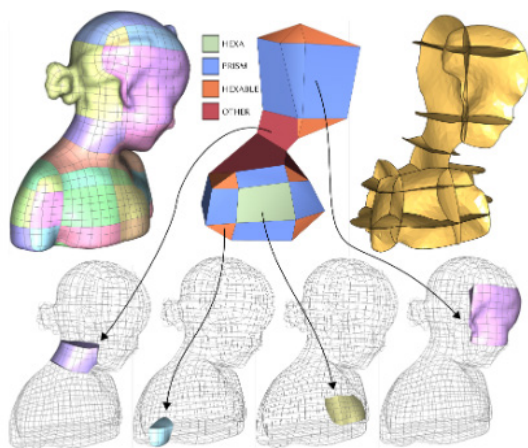ACM Transactions on Graphics, art. 121. ACM, 2020.

We present a new fully automatic block-decomposition algorithm for feature-pre-serving, strongly hex-dominant meshing, that yields results with a drastically larger percentage of hex elements than prior art. Our method is guided by a surface field that conforms to both surface curvature and feature lines, and exploits an ordered set of cutting loops that evenly cover the input surface, defining an arrangement of loops suitable for hex-element generation. We decompose the solid into coarse blocks by iteratively cutting it with surfaces bounded by these loops. The vast majority of the ob-tained blocks can be turned into hexahedral cells via simple midpoint subdivision. Our method produces pure hexahedral meshes in approximately 80% of the cases, and hex-dominant meshes with less than 2% non-hexahedral cells in the remaining cases. We demonstrate the robustness of our method
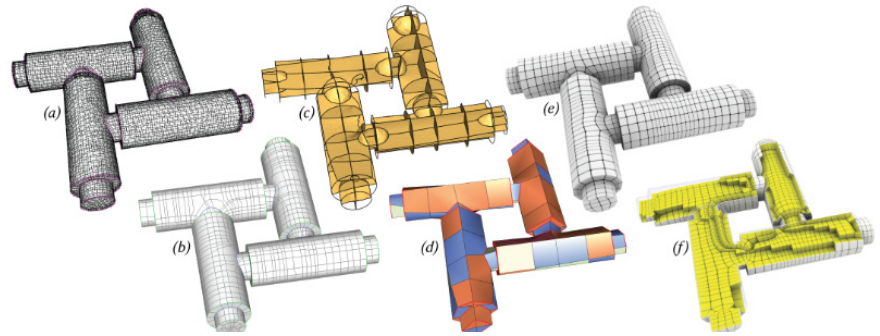
Given a surface mesh and a curvature and feature aligned cross-field (a) Loopy Cuts generates a sequence of field-aware cutting loops (b), and uses theseloops to generate solid cuts through the object (c), decomposing the model into a metamesh consisting of hex (green), prism (blue) and other (orange) simpleblocks (d). It converts the metamesh into a hex-mesh via midpoint refinement. The output hex-mesh (e,f) is well-shaped and well-aligned with the input field.

on 70+ models, including CAD objects with features of various complexity, organic and synthetic shapes, and provide extensive comparisons to prior art, demonstrating its superiority.

We define blocks as clusters of a labeled tetrahedral mesh (top left)bounded by cutting surfaces (top right). Each block defines an element of themetamesh (top, middle), and can be of four types: cuboids (green), prisms(violet), polyhedra with all valence three vertices that yield hexahedra undermidpoint refinement (orange), and other polyhedra which do not admit a pure hexahedral decomposition (red).

A cutting loop and its associated cut. Cuts are defined as cubicHRBF that interpolate loop points and the binormal (blue) to the surface(green) and loop tangent (red).

On improving the training of models for the semantic segmentation of Benthic communities from orthographic imagery

G. Pavoni, M. Corsini, M. Callieri, G. Fiameni, C. Edwards, P. Cignoni
Remote Sensing, vol. 12. MDPI, 2020.

The semantic segmentation of underwater imagery is an important step in the ecological analysis of coral habitats. To date, scientists produce fine-scale area annotations manually, an exceptionally time-consuming task that could be efficiently automatized by modern CNNs. This paper extends our previous work presented at the 3DUW'19 conference, outlining the workflow for the automated annotation of imagery from the first step of dataset preparation, to the last step of prediction reassembly. In particular, we propose an ecologically inspired strategy for an efficient dataset partition, an over-sampling methodology targeted on ortho-imagery, and a score fusion strategy. We also investigate the use of different loss functions in the optimization of a Deeplab V3+ model, to mitigate the class-imbalance problem and improve prediction accuracy on coral instance boundaries. The experimental results demonstrate the effectiveness of the ecologically inspired split in improving model performance, and quantify the advantages and limitations of the proposed over-sampling strategy. The extensive comparison of the loss functions gives numerous insights on the segmentation task; the Focal Tversky, typically used in the context of medical imaging (but not in remote sensing), results in the most convenient choice. By improving the accuracy of automated ortho image processing, the results presented here promise to meet the fundamental challenge of increasing the spatial and temporal scale of coral reef research, allowing researchers greater predictive ability to better manage coral reef resilience in the context of a changing environment.
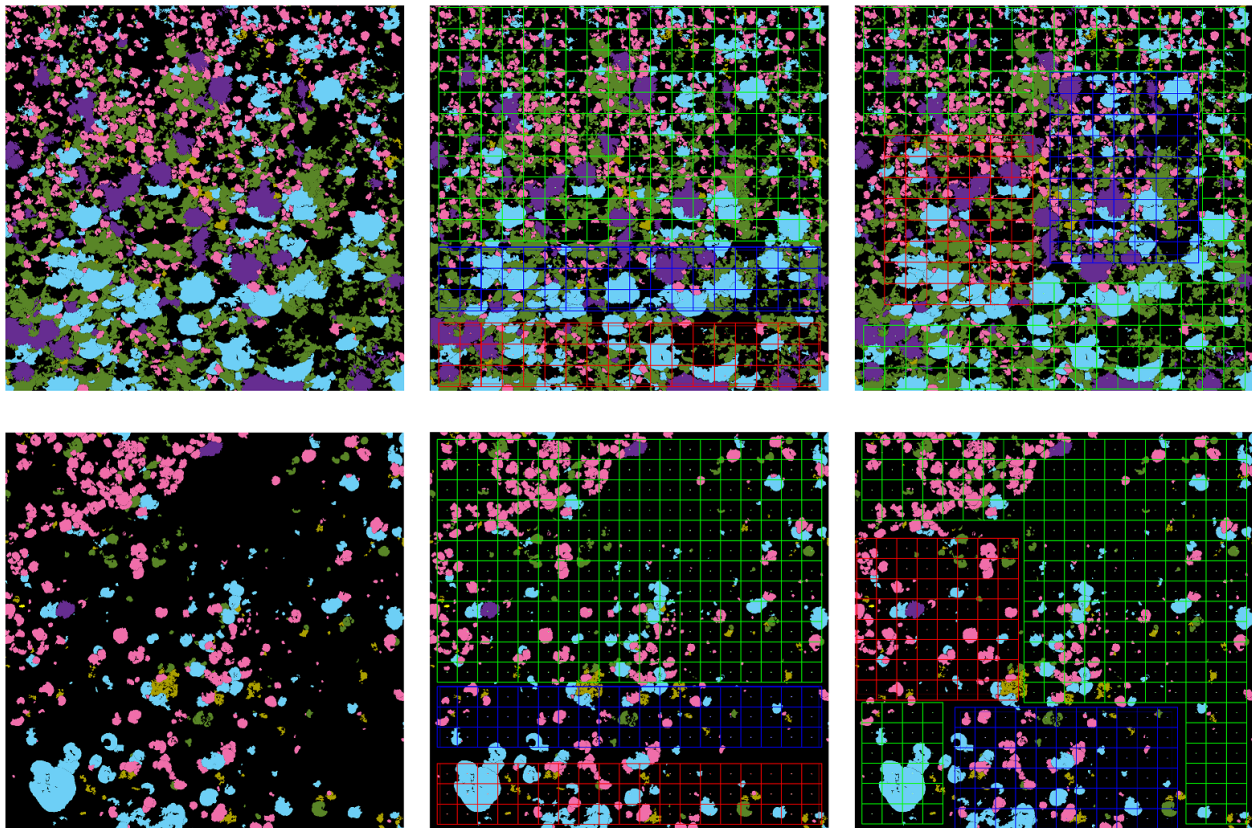
Results of the ecologically inspired partition of the orthos vs a uniform partition. The validation area is colored in blue, the test in red. The small rectangles represents the dataset clipped tiles. The first row is related to the labels of the FLI-1 plot, the second row to the MIL-4 plot.

A finite element model updating method based on global optimization

## M. Girardi, C. Padovani, D. Pellegrini, L. Robol
Mechanical Systems and Signal Processing, in press. Elsevier, 2020.

Finite element model updating of a structure made of linear elastic materials is based on the solution of a minimization problem. The goal is to find some unknown parameters of the finite element model (elastic moduli, mass densities, constraints and boundary conditions) that minimize an objective function which evaluates the discrepancy between experimental and numerical dynamic properties. The objective function depends nonlinearly on the parameters and may have multiple local minimum points. This paper presents a numerical method able to find a global minimum point and assess its reliability. The numerical method has been tested on two simulated examples – a masonry tower and a domed temple – and validated via a generic genetic algorithm and a global sensitivity analysis tool. A real case study monitored under operational conditions has also been addressed, and the structure's experimental modal properties have been used in the model updating procedure to estimate the mechanical properties of its constituent materials.
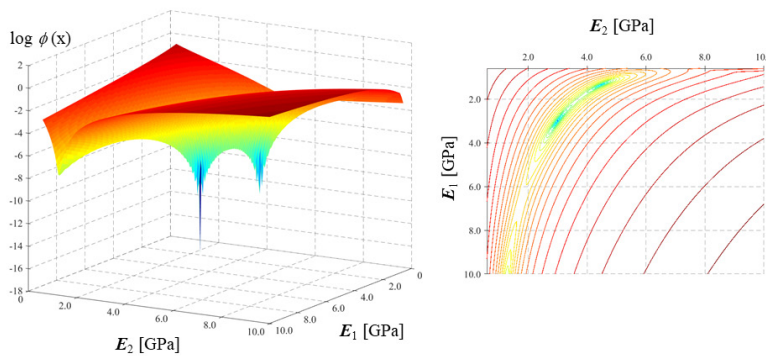
Case (a) – On the left a 3D plot of the objective function vs. and . On the right a contour plot of the same objective function where the two local minimum are clearly visible.

FE model of the Matilde donjon.
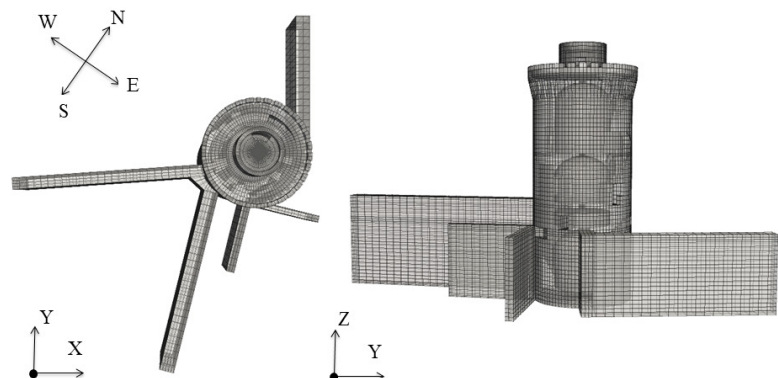
A 1% measurement of the gravitomagnetic field ofthe Earth with laser-tracked satellites

D. Lucchesi, M. Visco, R. Peron, M. Bassan, G. Pucacco, C. Pardini, L. Anselmo, C. Magnafico
Universe, vol. 6. MDPI, 2020.

A new measurement of the gravitomagnetic field of the Earth is presented. The measurement has been obtained through the careful evaluation of the Lense-Thirring (LT) precession on the combined orbits of three passive geodetic satellites, LAGEOS, LAGEOS II, and LARES, tracked by the Satellite Laser Ranging (SLR) technique. This general relativity precession, also known as frame-dragging, is a manifestation of spacetime curvature generated by mass-currents, a peculiarity of Einstein's theory of gravitation. The measurement stands out, compared to previous measurements in the same context, for its precision ($\approx 7.4 \times 10^{-3}$, at a 95% confidence level) and accuracy ($\approx 16 \times 10^{-3}$), i.e., for a reliable and robust evaluation of the systematic sources of error due to both gravitational and non-gravitational perturbations.

To achieve this measurement, we have largely exploited the results of the GRACE (Gravity Recovery And Climate Experiment) mission in order to significantly improve the description of the Earth's gravitational field, also modeling its dependence on time. In this way, we strongly reduced the systematic errors due to the uncertainty in the knowledge of the Earth even zonal harmonics and, at the same time, avoided a possible bias of the final result and, consequently, of the precision of the measurement, linked to a non-reliable handling of the unmodeled and mismodeled periodic effects.
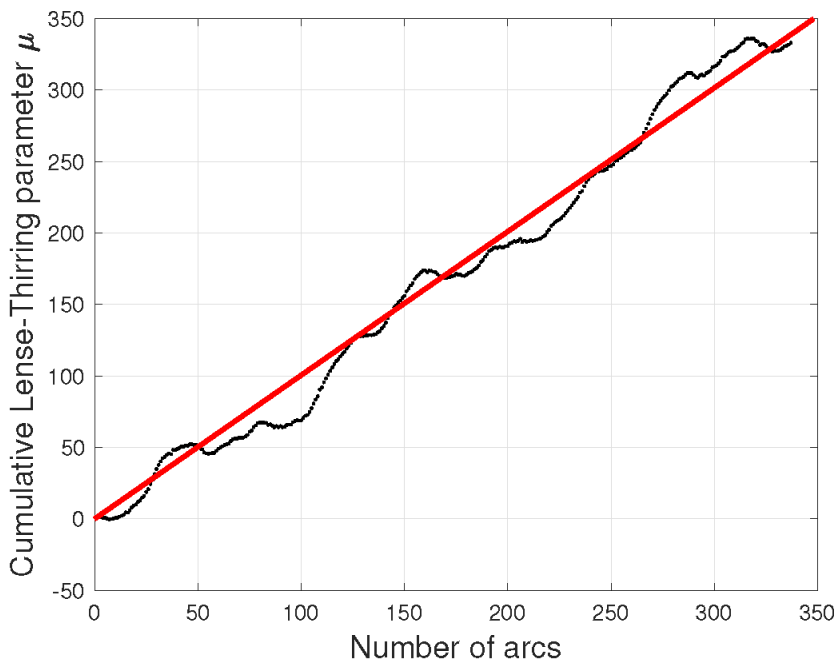
The plot shows the cumulative Lense-Thirring (L-T) parameter μ (black dots) as a function of the number of orbit determination arcs, of 7-day each, and for an overall timespan close to about 6.5 years. For each arc, the laser-ranged observations of the position of the satellites are compared with the corresponding position predicted by a dynamic model for the orbit of each satellite that takes into account the main gravitational and non-gravitational perturbations. A precise orbit determination, based on least squares fit of the tracking data, allows for the estimate of a corrected state-vector (i.e. position and velocity) for each satellite. The L-T parameter is then extracted by a separate analysis of the orbit residuals in the right-ascension of the ascending nodes of the three satellites. The slope of the red straight line obtained by a simple linear fit to the measured data corresponds to the value μ = 1.0053 ± 0074, where the error has a confidence level of 95%. The L-T parameter μ (with a value of 1 if Einstein's theory of General Relativity (GR) represents the correct theory for gravity and a value of 0 if Newtonian theory of gravitation is the correct one) represents a relativistic precession of the orbit of a satellite around the Earth due to the Earth's Gravitomagnetic field. The result obtained for μ agrees very well with the prediction of GR, with a formal discrepancy of about 0.7%, which rises to about 2% when the sources of systematic errors are also taken into account. This precession is a peculiarity of GR, where the Gravitomagnetic field is produced by the Earth's angular momentum, i.e. by the rotation of the Earth. The measurement of the precession confirms that mass-currents in the Universe contribute to the curvature of spacetime and can be considered as a weak manifestation of Mach's Principle.

SelfLens

Personal Assistive Technology to Support the Independence of People with Special Needs in Reading Information on Food Items

Selecting food items autonomously while shopping, and then storing and cooking these items can be a difficult task for people with special needs. Product labels on food packaging contain an ever-increasing amount of information, which can be in a variety of languages. The quantity of information and particular features of the text can make it difficult or impossible to read, particularly for the elderly or those with a visual impairment.

Online shopping can lead to further problems: web sites and applications may be inaccessible or difficult to use, interaction with web applications and the related technologies can be problematic, etc.. Even receiving and handling products at home (e.g., packets, boxes, etc.) can be a challenging task for people with special needs. Several tools or applications are available on the market or have been proposed to support this type of activity (e.g. barcode or QR code reading),

but they are limited and may require the user to have specific digital skills. For these reasons, difficulties when handling food items are often encountered by users with special needs.

In order to address this challenge, ISTI-CNR and Edi Group – a partner company in the food field – in collaboration with the Italian Association for the Blind in Tuscany (i.Ri.Fo.R. Regionale Toscano) have designed the SelfLens device: a simple portable tool which (1) can be used by anyone, regardless of their digital personal skills, (2) does not require a smartphone or complex device, (3) is a low-cost solution for the user, and (4) provides information both via audio feedback and text on screen.

SelfLens looks like a small remote control with a minimal user interface (two buttons, one display, and a speaker) that can be easily held in one hand and works simply by point-

ing at the product and pressing a button: the interaction is simple and practical for every user!

Using a specifically-designed code to mark any product and to overcome certain limitations in existing QR codes/barcodes, SelfLens assists in product recognition, information reading (e.g., ingredients), and product purchasing.

Edi Group now intends to set up a first experimentation of SelfLens in collaboration with a pilot supermarket in Italy.

Contact: Barbara Leporini, SI Lab
barbara.leporini@isti.cnr.it
http://www.selflens.com

Young researcher award "Matteo Dellepiane" 2020

The ISTI Young Researcher Award (YRA) "Matteo Dellepiane" is an annual award that honors its staff of less than 35 years old for a distinct contribute to the Institute activity with their scientific production.

There are two categories:

- Young Researcher - Beginner - awarded to researchers less than 32 years old;
- Young Researcher - Advanced - awarded to Ph.D. students and Ph.D. researchers less than 35 years old.

Contact: Franco Maria Nardini, HPC Lab
francomaria.nardini@isti.cnr.it

Beginner

Riccardo Guidotti

Francesco Laccone

Giulio Ermanno Pibiri

Advanced

Manlio Bacco

Luigi Malomo

Luca Pappalardo

Young Open Science award

Edition 2020

The ISTI Young Open Science Award aims at promoting the sharing and use of software and dataset produced during the research activities of the ISTI personnel. The award selects a publication connected to a software or dataset that has been made available to the community, and generated an impact in terms of uses, citations, visibility.

Recipient 2020:

- "A public data set of spatio-temporal match events in soccer competitions" by Luigi Pappalardo
- "A Validation Tool For Improving Semantic Segmentation of Complex Natural Structures" by Gaia Pavoni
- CDLIB: a python library to extract, compare and evaluate communities from complex networks by Giulio Rossetti

Contact: Franco Maria Nardini
francomaria.nardini@isti.cnr.it
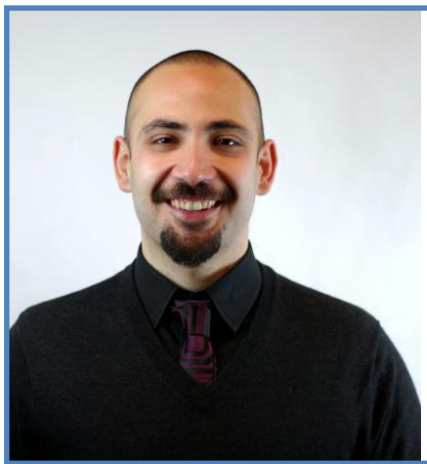
Luca Pappalardo

Gaia Pavoni

Giulio Rossetti

Best paper award at EG GCH 2020 conference

"Another Brick in the Wall: Improving the Assisted Semantic Segmentation of Masonry Walls" by Gaia Pavoni, Francesca Giuliani, Anna De Falco, Massimiliano Corsini, Federico Ponchio, Marco Callieri, and Paolo Cignoni presented at the XVIII Eurographics Workshop on Graphics and Cultural Heritage has received the best paper award at the conference.

In Architectural Heritage, the masonry's interpretation is an instrument for analyzing the construction phases, the assessment of structural properties, and the monitoring of its state of conservation.

This work is generally carried out by specialists that, based on visual observation and their knowledge, manually annotate ortho-images of the masonry generated by photogrammetric surveys.

This time-consuming manual work, often done with tools that have not been designed for this purpose, represents a bottleneck in the documentation and management workflow and is a severely limiting factor in monitoring large-scale monuments.

The work proposed in this paper explores the potential of AI-based solutions to improve the efficiency of masonry annotation in Architectural Heritage, experimenting the use of the TagLab open source tool (developed by the Visual Computing Lab), defining a workflow that support and empower the specialists' expertise.

http://vcg.isti.cnr.it/Publications/2020/PG-DCPCC20/

Three more gold medals at the RERS 2020 CTL parallel challenge

Franco Mazzanti (FMT lab), Frédéric Lang and Wendelin Serwe (CONVECS, Grenoble, France), after their successful participation in last year's RERS 2019 Challenge, again won three gold medals at the "Parallel CTL" track of RERS 2020. The RERS (Rigorous Examination of Reactive Systems) challenge is an international software verification competition held every year since 2011. RERS 2020 is the 10th edition of the challenge. Gold medals are awarded based on the number of problems solved, and the joint CNR-INRIA team solved 79 of the proposed 90 problems which were structured in three groups of increasing difficulty. Also this year, the result was obtained by the exploitation of advanced compositional verification techniques, some of which have been presented at the 3rd World Congress on Formal Methods (FM 2019) and implemented in the CADP toolbox developed at CONVECS.

http://rers-challenge.org/2020/

GreyNet International Award 2020

GreyNet International has conferred the GreyNet International Award 2020 to Silvia Giannini. The prize was awarded in recognition of her 15 years' contribution to the field of grey literature as an author and as a researcher, and as a special acknowledgement for the management of the Twenty-second International Conference on Grey Literature (GL2020) during a year challenged by the pandemic.

http://www.greynet.org/greynetmembership/annualaward.html

Pinocchio between art and technology

The Piazzetta dei Mosaici in Collodi comes alive thanks to augmented reality

There is an increasingly intense and evident fusion today between the concrete, material, habitable world and the virtual digital space where, at least potentially, everything is possible.

The VERO Project originates from these general premises and brings together the Carlo Collodi National Foundation and the Signal and Image Laboratory of ISTI. VERO is funded by the Tuscany Region. The project involves the creation of mobile applications that will animate one of the most important works of art in the Pinocchio Park, the Piazzetta dei Mosaici created by the artist Venturino Venturi.

"The activities of CNR began in June 2020 with the support of the Collodi Foundation and the collaboration of the Venturino Venturi Museum. With this project, another multimedia tool aimed at enhancing the work of Carlo Collodi and the monumental complex dedicated to the writer's literary masterpiece has been added to the Polycentric Park of Collodi." The augmented reality (AR) applications that are being developed will encourage the involvement of the visitors with an interactive audiovisual experience which can

The Fire-Eater (Venturino Venturi, La Piazzetta dei Mosaici).

be fun and informative at the same time. A version of the apps, dedicated to special AR see-through glasses, will offer a further degree of immersion, greatly enriching the experience of the visitors.

AR features could also be used together with printed material containing photographic reproductions of the mosaics (such as brochures and books), so tat the visitors can continue the interactive experience at

home, too. The virtual contents of the applications, created by professional digital artists, will be consistent with the magic of the artist Venturi and faithful to the spirit of Collodi's novel, thus preserving that idea of play and wonder of the book.

VERO began with a study of the Piazzetta and the style of its creator, Venturino Venturi. At the moment (October 2020), the development of the core AR technologies that will be used in the project has just begun. The project, which will end in spring 2022, is funded by the Region and falls within the POR FSE 2014-2020 (Regional Operational Program - European Social Fund).

Contact: Massimo Magrini, SI Lab
massimo.magrini@isti.cnr.itw

The Blue Fairy (Venturino Venturi, La Piazzetta dei Mosaici).

Ph.D. dissertations

Scalable processing and mining of big mobility data

Author: Omid Isfahani Alamdari, Dipartimento di Informatica, Università di Pisa
Supervisors: Dino Pedreschi, Roberto Trasarti

Spatial-temporal trajectory data contains rich information about moving objects and has been widely used for many real-world applications. However, the complexity of spatial-temporal trajectory data, on the one hand, and the fast collection of datasets, on the other hand, has made it challenging to efficiently store, process, and query such data. In this thesis, we proposed scalable methods to analyze big mobility data in the in-memory cluster computing environments where several considerations should be taken into account when partitioning, indexing, and processing these data. We developed a wide range of techniques and applications for large-scale spatial-temporal trajectory data management, aiming at scalable processing of such data. We addressed the novel research problem of the sub-trajectory similarity search query, in which, given a query trajectory, we try to find not only similar trajectories to the query but also sub-trajectories in data that are similar to the query (or sub-trajectories of the query). We used a two-level distributed indexing structure to support the search process and exploited a new technique for pruning unpromising parts of trajectories. We also proposed DISTANT framework for distributed annotation of vessel trajectories, which supports the segmentation of AIS trajectories based on the movement characteristics and annotating them using some predefined rules. The main focus of the framework is the identification of AIS switch-off gaps, as well as identifying the network of ports traveled by different ships. We developed these functionalities exploiting different indexing techniques and data processing units of Apache Spark.

Automatizing the large-scale analysis of underwater optical data

Author: Gaia Pavoni, Dipartimento di Ingegneria dell'Informazione, Università di Pisa
Supervisors: Luca Pollini, Andrea Caiti, Roberto Scopigno, Massimiliano Corsini

Underwater monitoring provides essential information to analyze the current condition and persisting trends of marine habitats. The optical data acquisition is a powerful solution to ensure both high-resolution and large-scale sampling of the seafloor. The use of autonomous data-driven robotics is making underwater imaging more and more popular.

Nevertheless, video and image sequences are trustworthy sources of knowledge that remain partially unexploited: the human visual analysis of images is a time-consuming task, which creates a bottleneck between data collection and extrapolation.

Nowadays, in visual identification tasks, the human effort could be efficiently automatized or supported by convolutional neural networks.

In this thesis, I discuss my experience with the automatic recognition of benthic species using deep learning-based methodologies, and I present TagLab, an AI-assisted software platform for the analysis of coral reefs.

Urban structure and mobility as spatio-temporal complex networks

Author: Gevorg Yeghikyan, Scuola Normale Superiore di Pisa
Supervisors: Mirco Nanni, Angelo Facchini, Marco Conti, Andrea Passarella, Bruno Lepri.

Contemporary urban life and functioning have become increasingly dependent on mobility. Having become an inherent constituent of urban dynamics, the role of urban moblity in influencing urban processes and morphology has increased dramatically. However, the relationship between urban mobility and spatial socio-economic structure has still not been thoroughly understood. This work will attempt to take a complex network theoretical approach to studying this intricate relationship throug:

- the spatio-temporal evolution of ad-hoc developed network centralities based on the Google PageRank;
- multilayer network regression with statistical random graphs respecting network structures for explaining urban mobility flows from urban socio-economic attributes;
- and Graph Neural Networks for predicting mobility flows to or from a specific location in the city.

Making both practical and theoretical contributions to urban science by offering methods for describing, monitoring, explaining, and predicting urban dynamics, this work will thus be aimed at providing a network theoretical framework for developing tools to facilitate better decision-making in urban planning and policy making.

Conferences - Co-organized by ISTI

http://www.aiucd.it/convegno-annuale/

Online Event March 28-April 1, 2021
https://www.ecir2021.eu/

http://fsen.ir/2021/

https://icc2021.ieee-icc.org/workshop/ws-27-work-shop-sage-green-solutions-smart-environment

http://ipin-conference.org/2021/index.html