BONSEYES: Platform for Open Development of Systems of Artificial Intelligence

Invited paper

Tim Llewellynn
nVISO SA
Switzerland
tim.llewellynn@nviso.ch

M. Milagro
Fernández-Carrobles
University of Castilla-La Mancha
Spain

Oscar Deniz
University of Castilla-La Mancha
Spain

Samuel Fricker
i4Ds Centre for Requirements
Engineering, FHNW
Switzerland

Amos Storkey
University of Edinburgh
United Kingdom

Nuria Pazos
Haute Ecole Specialisee de Suisse
Occidentale
Switzerland

Gordana Velikic
RT-RK
Serbia

Kirsten Leufgen
SCIPROM SARL
Switzerland

Rozenn Dahyot
Trinity College Dublin
Ireland

Sebastian Koller
Technical University Munich
Germany

Georgios Goumas
The Institute of Communications and
Computer Systems of the National
Technical University of Athens
Greece

Peter Leitner
SYNYO GmbH
Austria

Ganesh Dasika
ARM Ltd.
United Kingdom

Lei Wang
ZF Friedrichshafen AG
Germany

Kurt Tutschku
Blekinge Institute of Technology
(BTH)
Sweden

ABSTRACT

The Bonseyes EU H2020 collaborative project aims to develop a platform consisting of a Data Marketplace, a Deep Learning Toolbox, and Developer Reference Platforms for organizations wanting to adopt Artificial Intelligence. The project will be focused on using artificial intelligence in low power Internet of Things (IoT) devices ("edge computing"), embedded computing systems, and data center servers ("cloud computing"). It will bring about orders of magnitude improvements in efficiency, performance, reliability, security, and productivity in the design and programming of systems of artificial intelligence that incorporate Smart Cyber-Physical Systems (CPS). In addition, it will solve a causality problem for organizations who lack access to Data and Models. Its open software architecture will facilitate adoption of the whole concept on a wider scale. To evaluate the effectiveness, technical feasibility, and to quantify the real-world improvements in efficiency, security, performance, effort and cost of adding AI to products and services using the Bonseyes platform,

four complementary demonstrators will be built. Bonseyes platform capabilities are aimed at being aligned with the European FI-PPP activities and take advantage of its flagship project FIWARE. This paper provides a description of the project motivation, goals and preliminary work.

KEYWORDS

Data marketplace, Deep Learning, Internet of things, Smart Cyber-Physical Systems

ACM Reference format:

Tim Llewellynn, M. Milagro Fernández-Carrobles, Oscar Deniz, Samuel Fricker, Amos Storkey, Nuria Pazos, Gordana Velikic, Kirsten Leufgen, Rozenn Dahyot, Sebastian Koller, Georgios Goumas, Peter Leitner, Ganesh Dasika, Lei Wang, and Kurt Tutschku. 2017. BONSEYES: Platform for Open Development of Systems of Artificial Intelligence. In *Proceedings of CF'17, May 15-17, 2017, Siena, Italy, 2017*, 6 pages.
DOI: <http://dx.doi.org/10.1145/3075564.3076259>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CF'17, May 15-17, 2017, Siena, Italy

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4487-6/17/05.

DOI: <http://dx.doi.org/10.1145/3075564.3076259>

1 INTRODUCTION

Artificial intelligence (AI) is developing much faster than we thought and specially its emerging branch called deep learning, which is transforming AI. Deep learning relies on simulating large, multi-layered webs of virtual neurons. This enables a computer to learn

to recognize abstract patterns and solve any general pattern recognition problem. The results using the new methodology have been impressive, and both academia and the industry are currently moving at lightspeed towards the deep learning paradigm.

Thousands or millions of training examples are currently required by these state-of-the-art machine classification algorithms, while in contrast humans are able to learn from few examples. The current trend for improving AI to tackle increasingly complex problems is therefore a brute force solution to scale up the infrastructure: more data, more computing power, more neurons in deep learning algorithms, see for instance [2]. Tuning the very large number of latent parameters controlling the resulting deep architectures is a difficult optimisation problem where over-fitting (or learning the noise) is one of the major issues. More recent works aim at reducing the dimensionality of these architectures by imposing sparsity constraints in the latent space of parameters. The ability for the user to introduce any prior knowledge to help learning from small datasets is also essential for efficient architectures to be designed for a wide range of applications where data is scarce. Moreover learning efficiently from unstructured (or poorly structured) datasets and data streams that have noise is still currently a challenge for deep architectures.

Another problem arising from the increase of the deep learning techniques is related to GPUs. GPUs are regarded as one of the main drivers of deep learning. As such, the DNN (Deep Neural Network) models have, mostly, been designed with the GPU computational model in mind and may not be suitable for other platforms. On the other hand, the existing models have been mostly hand-crafted with no proof of optimality. Therefore, embedding these networks into other platforms may require either 1) developing new networks taking into account the specifications of the target platforms, or 2) adapting the existing models to the target platforms. This will have a direct impact on power efficiency, which is a key differentiator in the case of embedded (and wearable) platforms.

On the other hand, despite advances in data management and cloud infrastructure, AI systems are still predominantly developed as monolithic systems. A key reason is that offerings like Microsoft Azure Marketplace, Google Cloud Machine Learning, and IBM Watson require deployment of data into the clouds of the respective vendor and the use of the learning tools of these vendors. This approach makes a marketplace unattractive for data providers that are not willing to disclose their data, prevents scenarios in which data is produced by a network of devices (e.g. as in IoT), and impedes the emergence of a thriving ecosystem of data providers, algorithm developers and model trainers, and AI system developers. As a result, companies own the full AI systems development value chain instead of spreading the cost of ownership and accelerating time-to-market and quality through reuse of data and models.

Finally, the massive data collection required for deep learning presents obvious privacy issues [11]. Users' personal, highly sensitive data such as photos and voice recordings is kept indefinitely by the companies that collect it. Users can neither delete it, nor restrict the purposes for which it is used. Furthermore, centrally kept data is subject to legal subpoenas and extra-judicial surveillance. Many data owners for example, medical institutions that may want to apply deep learning methods to clinical records are prevented by

privacy and confidentiality concerns from sharing the data and thus benefiting from large-scale deep learning. Generally the best way to protect privacy is to confine it to the device so that sensor data never leaves it. This approach requires training of networks or adaptation of networks in the device.

In this context, the Bonseyes¹ collaborative project has been proposed and recently funded by the European Commission through its Horizon 2020 Research and Innovation Programme. In terms of scientific and technological background, the Bonseyes Project Consortium consists of leading researchers who have a strong mix of knowledge. This includes statistical and machine learning, embedded software and compiler optimization, power efficient hardware architectures, image processing and computer vision, cloud and distributed systems and software ecosystems and requirements engineering.

1.1 Objectives

The Bonseyes project aims to create a platform for open development of systems of AI, which is clearly emerging as a key growth driver in Smart CPS systems in the next decade. This is in contrast to monolithic system design currently used in closed end-to-end solutions. The main objectives of the project are summarized in Table 1.

2 METHODOLOGY

Bonseyes is a platform for open development in building, improving, and maintaining systems of AI in the age of the IoT. Figure 1 shows the problem that Bonseyes is solving, which is that Monolithic development of Systems of AI give rise to the Datawall effect that only large companies with end to end solutions can pass. In addition, Fig. 2 shows a summary of the target platforms that will be used in the project.

2.1 Data Marketplace

The objective of the Data Marketplace is to enable a modularized AI systems development value chain by offering the publishing, trade, and acquisition of data, metadata, and models. Data are measurements from the real world, typically made available as streams of data or as batches of archival data. Metadata enhances that data by capturing pre-processing results, classifying the primary data according to expert knowledge (e.g. for supervised learning), specifying the context the data relates to, and documenting feedback from AI system developers and users (e.g. for online learning). AI models are created by applying learning algorithms on data and metadata. The models embed knowledge about the data and classifications and enable automated classification. All three components are used by AI system developers as components to build smart systems.

2.2 Deep Learning Toolbox

The objective of the Deep Learning Toolbox is to provide a set of deep learning components that are tailored for embedded, constrained, distributed systems operating in real environments with noisy, sometimes missing data. The toolbox will enable the selection

¹www.bonseyes.com

Figure 1: The problem that Bonseyes is solving.

Figure 2: Target platforms.

and optimisation of tools for a particular task. The key components of the toolbox are:

2.2.1 Deep learning methods. Methods that are flexible in terms of representation: number of bits precision, hashing methods, parameter sharing structure, memory usage, sparsity, and robustness.

Table 1: Bonseyes objectives

Objectives
Accelerate the design and programming of systems of artificial intelligence. Design and implement a Data Marketplace. Reusing of data, meta data, and models among separate legal entities to reduce design and development time in building systems of AI as compared to existing monolithic system of systems design and development approaches.
Reduce the complexity in training deep learning models for distributed embedded systems. Provide the fundamental tools for deep learning for constrained architectures. Design and implement noise resistant machine learning capabilities able to learn efficiently from unstructured, partially labelled datasets. Design and implement sparse models to scale and adapt to various computational architectures. Enable users for designing deep learning models taking advantage of prior domain knowledge information.
Foster "embedded intelligence" on low power and resource constrained Smart CPS. Bonseyes will enable the development of deep learning models tailored for low power embedded systems by incorporating architecture-awareness in existing standardised deep learning packages via an open source toolbox. Deep Learning Toolbox will allow deep models to be designed and efficiently deployed across multiple embedded systems and support multiple low power developer reference platforms.
Demonstrate the whole concept in four key challenging scenarios. Demonstrate the technical feasibility of Bonseyes with at least three Developer Platforms in four scenarios: automotive safety, automotive personalisation, consumer devices, health-care. Each scenario will involve the creation of a specific application using the Data Marketplace together with one or more Developer Platforms to build systems of AI

2.2.2 Cost sensitive optimisation methods. Methods (e.g. Bayesian optimisation) that can optimise a particular deep learning architecture for a particular embedded environment by incorporating the particular individual costs (e.g. cost of using memory resources, cost of processing power, energy cost, training time costs etc.) of that environmental system into the optimisation. This includes obtaining reduced dimensional parametrisations for particular architecture classes, rather than just individual architectures. We will explicitly encode for the costs of on-device learning, and costs (e.g. communication costs) for cloud-based learning.

2.2.3 Component valuation. The value of a given deep learning component in a given environment will be considered an integral part of the component itself. The Deep Learning Toolbox will also include a measure of when a component would be useful. The information value of both data and deep-learning components can then provide a structural marketplace for deciding what components are best used with what data.

2.2.4 Transferability metrics. A component can either be a free form deep learning model that needs optimizing for a particular data source, or a learnt component, already optimised on different data source, or cost function.

2.2.5 Structure-sensitive implementation. Often generic low-level linear algebra libraries are used for implementing deep neural networks, but such libraries rarely provide the best computation for the restricted architectures needed for embedded systems. For example, both structural sparsity and dynamic sparsity (associated with the use of sparse activation functions) provide substantial opportunity for computational saving.

2.2.6 Implementation and tailoring of these components for particular reference architectures. The targets across all components are the real systems that the deep learning components need to work on. All development will use specific embedded environments as exemplars. The specific development of deep learning methods for the individual environments will form a substantial part of the development effort.

2.3 Universal Developer Reference Platforms

The Deep Learning Toolbox will provide a unified framework for accelerating Deep Convolutional Neural Networks on resource constrained architecture-dependent embedded systems. A number of platforms will be made available with pre-integrated middle-ware via open source packages. This offers developers a number of advantages:

- **Heterogeneous:** Supports a wide array of CPU-based platforms, VPUs, and DSPs.
- **Performance:** Optimised code for reduced memory, power, and CPU overhead.
- **Scalability:** Ability to run models on the cloud or on embedded systems.
- **Configurability:** Support for multiple types of deep neural network architectures.
- **Concurrent Classification & Learning:** Support for incremental learning at runtime through feedback APIs, which allows for more flexible and general network.

The following platforms will be supported by consortium partners ARM, RT-RK, and HES-SO:

- ARM-based platforms are used extensively when deploying artificial intelligence in embedded and automotive environments. Optimizing the deep learning toolbox for an ARM-based platform is a clear choice.
- Low power CPU/VPU/GPU-based platforms are emerging for always-on visual intelligence applications based on low power vision processors. Standing at the intersection of low-power and high performance, these platforms enable the development of embedded intelligent solutions. Equipped with multiple sensors, cameras and communication means, they will be used during the development and validation stages of the networks prior to deployment and integration.
- DSP (Digital Signal Processor)-based platforms provided by RT-RK have been successfully deployed for a spectrum of consumer electronics applications, complex industrial settings, and highly demanding automotive environments and military applications. The board will support up to ten cameras, basic and advanced warning systems, active control systems and semi-autonomous applications.

Apart from the platforms above, Android smartphones will be also considered in the Bonseyes project.

2.4 Systems of artificial intelligence

Most AI systems involve some sort of integrated technologies, for example the integration of speech synthesis technologies with that of speech recognition. However, in recent years there has been an increasing discussion on the importance of systems integration as a field in its own right. AI integration has recently been attracting attention because a number of (relatively) simple AI systems for specific problem domains (such as computer vision, speech synthesis, etc.) have already been created. Integrating what is already available is a more logical approach to broader AI than building monolithic systems from scratch. Within Bonseyes, four demonstrators will be used to build such systems of AI (c.f. section 2.7).

2.5 Computing Power

The objective of the Computing Power component is to provide resources in terms of CPU, memory, and storage to provide the capabilities for the Data Marketplace. Bonseyes will use a well-balanced approach for providing the training in the cloud backend with sufficient computing power. The balance aims at economic and energy efficiency, robustness, functional capabilities and data privacy and isolation. It will mainly be based on the use of the EU flagship platform FIWARE². The Bonseyes project will use a balanced and agile concept of non-commercial, commercial and, if necessary, self-operated compute infrastructure sourced from FIWARE LAB³. These options include possible special support by FIWARE Lab. The balance will reflect on the economics, availability and stability of the compute resources and the needs of the use cases. The consideration of available resources will increase the agility of the Bonseyes concept of a system of AI systems. The detailed balance will be determined during the architecting phase of the Bonseyes project.

2.6 Data Tools

Data Tools aims to provide tools to allow data collection, curation, and augmentation: downloading, uploading, versioning, labelling, evaluating, crowdsourcing, and editing data necessary for training models using the Deep Learning Toolbox. One key area will be on IoT data collection by providing a programming model and micro-kernel style runtime. That can be embedded in gateways and small footprint edge devices enabling local, real-time, analytics on the continuous streams of data coming from equipment, vehicles, systems, appliances, devices and sensors of all kinds. By performing real-time analytics on the edge device, only anomalies or unseen data can be transmitted for storage and archival used for learning.

2.7 Demonstrators

For demonstration, four scenarios in three sectors (automotive, consumer, and healthcare) will be considered: automotive intelligent safety, automotive cognitive computing, consumer emotional virtual agents, and healthcare patient monitoring. These use cases

have been considered as they are far reaching across a number of high-value industries with high social impact.

2.7.1 Automotive Intelligent Safety. Autonomous systems are able to control steering, braking, and accelerating and are already starting to appear in cars. These systems require drivers to keep an eye on the road and hands on the wheel. But the next generation of self-driving systems [4], [8] could be available in less than a decade and free drivers so they can work, text, or just relax. Ford, General Motors, Toyota, Nissan, Volvo, and Audi have all shown off cars that can drive themselves. They all have declared that within a decade they plan to sell cars with some form of advanced automation. These cars will be able to take over driving on highways or to park themselves in a garage.

In this demonstrator, Bonseyes will be used to build a system of AI that will use scene and people detection (contextual awareness) and driver distraction (driver monitoring) to trigger active or passive safety systems. It will involve the following AI systems: per-pixel scene labelling, scene detection, people detection, and driver distraction.

2.7.2 Automotive Cognitive Computing. The vehicles of the near future will be "intelligent". Electronics will bring new capabilities to every part of the vehicle. New technologies will provide for greater assistance in navigation, enhanced driver information about the vehicle, its environment [10] and vehicle connectivity. Consumers, with a plethora of electronic devices that inform them [5], entertain them and keep them safe [9], [6], will find themselves enjoying the overall experience of their vehicles. Connectivity and lifestyle trends will change the way cars are used. This "experience" will be a key differentiator in attracting consumers.

In the Automotive Cognitive Computing Demonstrator, Bonseyes will be used to build an in-vehicle digital assistant. The in-vehicle digital assistant will be able to recognise the driver and then personalise his/her car experience while learning the driver's preferences and allowing natural language interaction within the vehicle and the driving context. It will provide the driver with personalised advice on how to interact with the vehicle for route planning, environment information, entertainment, etc. It will involve the following AI systems: face recognition, demographic detection, emotion recognition, speech-to-text, and natural language processing.

2.7.3 Consumer Emotional Virtual Assistants. Many institutions are creating innovation labs aimed at understanding how they can rewrite their existing applications to improve the technology and consumer interaction and to provide a more compelling customer service experience. Human sensor data, such as facial expressions [12], voice input, hand gestures, even brain waves, emotion sensors, heart beat are being tested as new forms of input [1]. While voice response systems, haptics (tactile feedback) and holographs are being tested as new forms of output. These inputs and outputs are being augmented with AI to make them more useful and human-like enabling, for example, discourse in natural language and sensing emotions. Based upon what is learned, a new paradigm will likely emerge that will fundamentally change the way machines and people communicate.

Bonseyes will be used to leverage the increasing amount of computational capacity of mobile devices to develop real time multimodal

²www.fiware.org

³lab.fiware.org

applications. An emotional virtual agent will be implemented for improving the communication between services and users through an agent-based application that allows multimodal and emotional interaction with the users through different channels: visual, oral and written. A use case will be developed showing how multiple sensing technologies can be combined with object detection to enhance consumer interaction in a more natural way. It will involve the following AI systems: face recognition, multi-modal emotion recognition, object recognition, and speech-to-text.

2.7.4 Healthcare Patient Monitoring. Patient tracking will be used to optimise capacity utilisation in diagnostic departments and reduce waiting times for patients. The vital sensor delivers data like acceleration, pose, heart rate and others which will be used to estimate the mobility of the patient and the time needed to reach the diagnostic department [3], [7]. Personal recording of vital parameters and the use of health apps is already widespread not only in Europe. Hospitals started in the last years to monitor patient beds, devices and patients by the use of RFID, however, with the significant drawback of limited range and necessary huge installations for the RFID antennas in the corresponding areas. In the future, it will be focused on patient tracking and monitoring of different vital signals by use of smart low power devices in order to plan and schedule further diagnostic procedures and calculate expected stress and therapeutic options by use of deep learning techniques. Patients scheduled for elective surgery will be equipped during the diagnosis day with smart devices which track their position and record and transmit vital signs (heart rate, breathing rate, pose, skin conduction, etc.). Based on necessary diagnostics patients will be sent to the respective diagnostic department. The Deep Learning Toolbox will analyse vital signs data and predict further diagnostics and stress levels which will be used to adjust sedation prior to the surgical intervention. Postoperatively, vital data from the sensors will be used to predict the earliest day of discharge. It will involve the following AI systems: vital sensors, and location tracking technologies.

3 CONCLUSIONS

The main challenge and contribution of the Bonseyes collaborative project is to design and implement highly distributed and connected digital technologies that are embedded in a multitude of increasingly autonomous physical systems. These systems must satisfy multiple critical constraints including safety, security, power efficiency, high performance, size and cost. Developing new model-centric and predictive engineering methods and tools for CPS with a high degree of autonomy, ensuring adaptability, scalability, complexity management, security and safety providing trust to humans in the loop. Driven by industrial needs and validated in at least four complementary use cases in different application domains and sectors. The results are intended to enable integration of broader development environments and middleware. The merits of the contribution should be made explicit.

4 ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732204 (Bonseyes). This work is supported by the Swiss

State Secretariat for Education Research and Innovation (SERI) under contract number 16.0159. The opinions expressed and arguments employed herein do not necessarily reflect the official views of these funding bodies.

REFERENCES

- [1] Sourav Bhattacharya and Nicholas D. Lane. 2016. From smart to deep: Robust activity recognition on smartwatches using deep learning. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. 1–6. DOI : <http://dx.doi.org/10.1109/PERCOMW.2016.7457169>
- [2] Xue-Wen Chen and Xiaotong Lin. 2014. Big Data Deep Learning: Challenges and Perspectives. *IEEE Access* 2 (2014), 514–525. DOI : <http://dx.doi.org/10.1109/ACCESS.2014.2325029>
- [3] Shin Hoo-Chang, Orton Matthew R., Collins David J., Doran Simon J., and Leach Martin O. 2013. Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 35, 8 (2013), 1930–1943.
- [4] Brody Huval, Tao Wang, Sameep Tandon, Jeff Kiske, Will Song, Joel Pazhayampallil, Mykhaylo Andriluka, Pranav Rajpurkar, Toki Migimatsu, Royce Cheng-Yue, Fernando Mujica, Adam Coates, and Andrew Y. Ng. 2015. An Empirical Evaluation of Deep Learning on Highway Driving. *CoRR* abs/1504.01716 (2015).
- [5] Ashesh Jain, Hemma S. Koppula, Shane Soh, Bharad Raghavan, Avi Singh, and Ashutosh Saxena. 2016. Brain4Cars: Car That Knows Before You Do via Sensory-Fusion Deep Learning Architecture. *CoRR* abs/1601.00740 (2016). <http://arxiv.org/abs/1601.00740>
- [6] Ashesh Jain, Avi Singh, Hemma S. Koppula, Shane Soh, and Ashutosh Saxena. 2015. Recurrent Neural Networks for Driver Activity Anticipation via Sensory-Fusion Architecture. *CoRR* abs/1509.05016 (2015). <http://arxiv.org/abs/1509.05016>
- [7] Guo-Ping Liu, Jianjun Yan, Yiqin Wang, Wu Zheng, Tao Zhong, Xiong Lu, and Peng Qian. 2014. Deep Learning Based Syndrome Diagnosis of Chronic Gastritis. *Comp. Math. Methods in Medicine* 2014 (2014), 938350:1–938350:8. DOI : <http://dx.doi.org/10.1155/2014/938350>
- [8] Qudsia Memon, Muzamil Ahmed, Shahzeb Ali, Azam R. Memon, and Wajiha Shah. 2016. Self-driving and driver relaxing vehicle. (Nov 2016), 170–174. DOI : <http://dx.doi.org/10.1109/ICRAI.2016.7791248>
- [9] Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Kari Pulli. 2015. Multi-sensor system for driver's hand-gesture recognition. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 1. 1–8. DOI : <http://dx.doi.org/10.1109/FG.2015.7163132>
- [10] David Ribeiro, André Mateus, Jacinto C. Nascimento, and Pedro Miraldo. 2016. A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation. *CoRR* abs/1607.04441 (2016). <http://arxiv.org/abs/1607.04441>
- [11] Reza Shokri and Vitaly Shmatikov. 2015. Privacy-Preserving Deep Learning. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. ACM, New York, NY, USA, 1310–1321. DOI : <http://dx.doi.org/10.1145/2810103.2813687>
- [12] Inchul Song, Hyun-Jun Kim, and Paul B. Jeon. 2014. Deep learning for real-time robust facial expression recognition on a smartphone. In *2014 IEEE International Conference on Consumer Electronics (ICCE)*. 564–567. DOI : <http://dx.doi.org/10.1109/ICCE.2014.6776135>