March 2020

DEEL
DEpendable & Explainable Learning

DEEL MAG #03

TOULOUSE & QUEBEC JOIN FORCES TO DEVELOP AI FOR CRITICAL SYSTEMS

IRT SAINT EXUPERY REMAINS MOBILIZED

Given the current situation in France, and in order to comply with government decisions as far as possible, the **IRT Saint Exupery took the necessary measures to ensure the safety of its staff and the continuity of its activities**.

Access to our sites in Bordeaux (ENSAM Campus), Montpellier, Montreal, Sophia-Antipolis (INRIA) and Toulouse (B612) is now limited.

Teleworking has been set up for all the employees of IRT Saint Exupery. Special arrangements have been made for employees whose activity does not allow them to make the transition to teleworking.

IRT Saint Exupery has suspended all its participation in external events and also postponed all the events we were organizing. Click Here for More Information

Do not hesitate to get in touch with your usual interlocutors, who will remain present and available throughout this period.

#COVID-19 – STAY HOME – STAY SAFE

AI SAFETY LANDSCAPE, SAFEAI AND AAAI-2020 CONFERENCE

From February 6th to February 12th, DEEL (through Grégory FLANDIN, Mélanie DUCOFFE and Jayant SEN GUPTA) was represented at three gatherings dealing with AI and safety organized in New York.

AI safety landscape initiative gathers, on invitation for the moment, key players in AI safety from academia, governmental organizations and industries, mostly from transport, to draw the big picture of this new domain, from design to life-long learning and including Artificial General Intelligence (AGI). **DEEL was identified as a major contributor in this field and pushed for a focus on narrow AI, which is already challenging to certify.**

SafeAI is a series of workshops associated this year to AAAI conference for its second edition. **The paper on probabilistic guarantees for shifted surrogate models was presented and selected among best papers candidates.**

Finally, in AAAI, already impacted by COVID-19 outbreak in China, 25% of the presentations were done by videos to cope with Chinese researchers' impossibility to present in the conference. Highlights of the conference were really great plenary conferences that were recorded and videos are available here: https://aaai.org/Conferences/AAAI-20/livestreamed-talks/

# Jayant SEN GUPTA, Mélanie DUCOFFE, Grégory FLANDIN

LIPSCHITZ NEURAL NETWORK LIBRARY

Rudolf Lipschitz

Robustness of Neural Network to adversarial samples [1] is an important challenge, and ***k-lipschitz*** property for Neural Network was pointed out as an intrinsic positive factor for robustness. A function f is said *k-Lipschitz* when its first derivatives are bounded by k, or equivalently when function outputs distance $||f(x+\varepsilon)-f(x)||$ is lower than k times the distance of its inputs k.$||\varepsilon||$, certifying that variations on outputs are bounded within a neighborhood of. This *k-lipschitz* property, besides local robustness, is also mandatory for Wasserstein distance estimation for instance in WGAN [2]. If computing the real *k* factor for a full Neural network is *NP-complex*, several solutions exist to control the factor of each layer (spectral normalization, L2-normalization, orthonormalization …) and activations [3]. DEEL core team has developed a **library** (based on tensorflow-2) **to construct and learn *k-lipschitz* convolutional neural networks with guaranties for each layer, and to export weights into conventional layers for inference after learning. This library is already available for DEEL industrial partners**, and we plan to publish it on Github. Applications for robust classification are under study.

[1] C. Szegedy et al. « Intriguing properties of neural networks », ICLR 2014

[2] M. Arjovsky et al. « Wasserstein Generative Adversarial Networks », ICML 2017

[3] C. Anil et al. « Sorting out Lipschitz function approximation », ICML 2019

# Thibaut BOISSIN, Mathieu SERRURIER, Franck MAMALET

FOCUS ON A PhD STUDENT, David BERTOIN

My profile is a little bit atypical: after a first diploma in Mathematical Engineering obtained in 2012 at the University of Paris VI Pierre and Marie Curie, I started my career as a research engineer in operational research in Barcelona. I quickly returned to Paris to join a large group specialized in data processing and then a start-up specialized in recommendation engines. After a few years, I wanted to orient myself towards research and made the choice go back to school by integrating the Master Data-Science offered by the establishments' members of the current Polytechnic Institute of Paris.

After this second diploma, I decided to join the IRT Saint-Exupery and more specifically the DEEL project in 2018 as a research engineer. After integrating the project, Grégory FLANDIN and Guillaume GAUDRON gave me the opportunity to do a **thesis in reinforcement learning**.

I started my thesis on October 2019 under the supervision of Emmanuel RACHELSON and Sébastien GERCHINOVITZ. The subject of my thesis focuses **on the interest of describing states through objects in reinforcement learning**. Inspired by the framework introduced by Carlos Diuk in "An object-oriented representation for efficient reinforcement learning ", we seek to benefit from a high-level representation of the state space through an object-oriented description **to gain not only on interpretability but also on transfer during the learning of several tasks**. For the moment, I am particularly interested in the problem of the Sokoban which lends itself particularly to the description in object and whose different configurations of labyrinths correspond to different tasks in which the transitions between states although different obey the same physical laws.

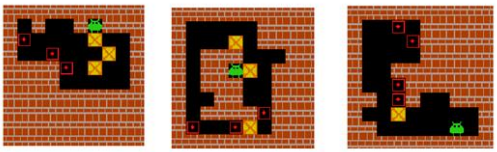

Figure : Different tasks in Sokoban

# David BERTOIN, Sébastien GERCHINOVITZ, Emmanuel RACHELSON

KEY DATES

26 & 27 Mar	Certification mission Workshop	26 & 27 May	Mobilit.AI is postponed to 29th & 30th September
23 & 24 Apr	Certification mission Workshop	28 May	DEEL - International Operational Committee Cancelled ➢ Another meeting will be proposed