Preparation of an International Competition on the Distillation of Deep Networks learned on Sequential Data

Advisors: Rémi Eyraud, Antoine Bonnefoy
Mail: `Remi.Eyraud@univ-st-etienne.fr`
Locations: Hubert Curien Laboratory, Campus Manufacture, Saint-Etienne; EURA NOVA R&D Center, downtown Marseilles
Team: Data Intelligence
Level: Master 2
Gratuity: 3,90€ per hour (≈ 550€ per month) + cost of missions in Marseille
Associated PhD thesis opportunity: no

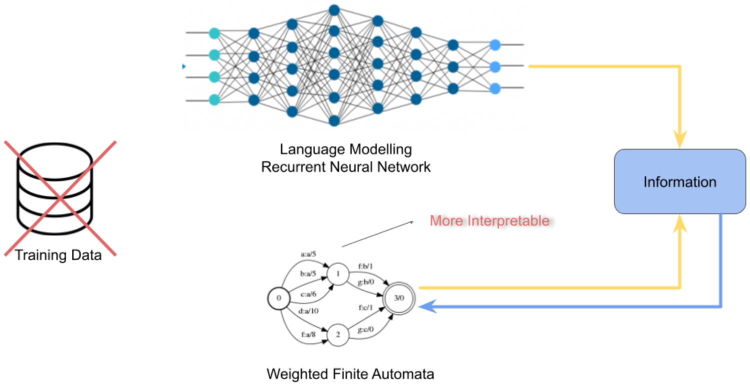

Figure 1: Illustration of the Knowledge Distillation principle

Summary. Recent successes of Machine Learning (ML), in particular the deep learning approach, and their growing impact on numerous fields have risen questions about the induced decision process. Indeed, the most efficient models are often overly parameterized black boxes whose inner ruling system is not accessible to human comprehension. However, to be able to understand how these models works is a crucial issue for the future developments of ML: this knowledge is a needed element for the development of the field, when it is not a legal requirement (*e.g.*, the RGPD).

One way to tackle this problem is what is called **Knowledge Distillation** [1], a framework that aims at extracting simpler, more efficient models from already learned, complex, uninterpretable ones (huge deep architectures, ensemble models, ...). This is usually presented as a teacher-student schema, where the teacher/black-box provides information to the student.

In the usual distillation framework the student is a smaller NN, as illustrated in Figure 1. But several recent works focus on the case where the student model offers an increased degree of interpretability, like Finite State Automata, Context-Free Grammars, or Tree formalism's (see [3, 2] for recent examples).

The **objective of this internship** is to set up an international competition where a set of neural networks already trained on sequential data will be supplied to the competitors. Their objective will be the extraction of a simpler model which exhibits performances similar to the original deep model.

A large panel of deep architectures should be covered, from simple RNN to attention-based NN like LTSM, GRU and Transformers [4]. The type of data used to train these models should also cover a large spectrum of applications from Natural Language Processing to audio or video signals. This will allow to tackle problems such as eye tracking, sentiment analysis or speech analysis. At the end of the competition, the goal is to have a Benchmark concerning a large number of possible use-cases so that researchers will be able to compare their new Distillation ideas on a state-of-the-art dataset. The competition should also provide baseline algorithms so that competitors have access to examples of code that they can work from if they want.

In addition to the creation of the data, which will require a study of available learning datasets and deep architectures for sequential data, the research work will address the questions of the quantitative and qualitative evaluation of distilled models, the comparison of existing distillation algorithms, and potentially the design of a new distillation algorithm.

Short stays in Marseilles are scheduled during the internship as the start-up EURA NOVA is involved in the construction of the competition and thus in this internship.

Expected results An important amount of objectives can be drawn to prepare this competition. The intern will have to chose among which:

- The study and preparation of sequential learning data
- The creation of 40 to 60 trained deep models of various architectures on these data
- The study of existing evaluation metrics for Knowledge Distillation and potentially the creation of new one(s) based on the simplicity and/or the interpretability of extracted models.
- The development of easy to use baseline(s) or the simplification of existing code to provide convenient code
- The study of the behavior of the baseline(s) on the trained deep models

- The selection of 15 to 20 trained models in order to have models of various complexity to distill during the competition
- The design of the on-line competition interface

Keywords: Recurrent Neural Network ; On-line competition ; Knowledge Distillation

[1] N. Frosst and G. E. Hinton. *Distilling a Neural Network Into a Soft Decision T*ree. In:W. on Comprehensibility and Explanation in AI and ML. 2017

[2] R. Eyraud and S. Ayache. *Distillation of Weighted Automata from Recurrent Neural Networks using a Spectral Approach*, Machine Learning Journal. 2021

[3] G. Weiss, Y. Goldberg, E. Yahav. *Extracting Automata from Recurrent Neural Networks Using Queries and Counterexamples.* In:ICML. 2018

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* In:NA-ACL. 2019