Machine Learning for Sequential Interactive Models

Supervisor: Pr. Ludovic DENOYER - UMR 7606 Laboratoire d'informatique de Paris 6

Context

Recently, a new family of algorithms called Sequential Prediction Models has emerged in Machine Learning. The motivation behind their development is that, while there is a plethora of standard algorithmic solutions for supervised and unsupervised training, most of the methods learn a monolithic predictor function, that is each test instance is processed in a single-step, atomic process. In contrast, some recent studies have proposed a different paradigm in which prediction is reformulated as a sequential decision process, and for which learning the predictor function corresponds to solving a dynamic control problem. Both the motivations and applications of these studies are quite diverse. Unrolling monolithic functions into a sequence of decisions and controlling when the final prediction is made is obviously beneficial to learn "budgeted predictors", so some of the approaches are motivated by either fast prediction or by keeping the number of costly feature evaluations low [1, 9]. Typical applications of these techniques are real-time object detection or web page ranking, trigger design in high-energy physics, or resource-bounded information extraction [7]. Despite the promising applications and diverse algorithmic solutions, sequential prediction still remains a young and unexplored field with numerous unanswered theoretical and algorithmic questions. A very interesting point is that sequential learning models can naturally take into account an explicit interaction with an expert, at two different levels :

- while sequential models behave "like experts", humans can be use to guide the learning process, and
- sequential models can naturally interact with humans by asking questions, as it has been recently done for the cold-start recommendation setting in [2].

Objectives

The Thesis will be organized among two different research axis focused on the two points described previously.

Attention-based/Transformation-based models

Recently, some sequential models has been proposed that are able to focus their attention on relevant sub-parts of any given input datum. These models have been mainly applied on text categorization [4] where they focus on relevant sentences of a document, on vectorial input data [5] where they act as adaptive features selection models, and on image classification [8, 6] where they simulate "eyes trajectories", focusing on relevant parts of the picture. Moreover, with a close formalism [3], variants of sequential models have been proposed that can decide how to transform inputs i.e which sequence of processing to apply on a particular datum. The image classification is a typical use-case, where pictures are first normalized (light, size, contrast, etc.) before being provided to an automatic system. The first contribution of the thesis will be to unified these two approaches (attention and transformation) in a unique model able to both learn where to get relevant information, but also how to transform collected information for a particular problem. While this problem will involve many different possible "actions", the work will also focus on new ways to train such models, and particularly by using imitation techniques: the learning process can be efficiently guided by humans or experts that will provide to the system which sequence(s) of actions is/are the most relevant for predicting. The system will thus benefit from a human-interaction to improve both its learning speed, but also its general quality.

Interactive sequential models

At last, while the previous axis focuses on the development of automatic predictive systems, the interaction with humans occurring during learning, we will also focus on extending sequential predictive models to the case where, at each time-step, the model can ask "questions" to experts. In other words, interacting with an expert will be one of the possible actions the model can decide to do at each time-step. By querying humans, the model will be able to collect high-level and high-quality information that will help him to predict over complex and difficult

inputs where classical systems fail. The developed systems will thus be an interactive system that will collect information from both information sources and humans, the collected data being then used as new training data.

References

[1] Robert Busa-Fekete, Djalel Benbouzid, and Balazs Kegl. Fast classification using sparse decision dags. In Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012, 2012.

[2] Gabriella Contardo, Ludovic Denoyer, and Thierry Artieres. Representation learning for cold-start recommendation. CoRR, abs/1412.7156, 2014.

[3] Ludovic Denoyer and Patrick Gallinari. Deep sequential neural network. arXiv preprint arXiv:1410.0510 - Workshop Deep Learning NIPS 2014, 2014.

[4] Gabriel Dulac-Arnold, Ludovic Denoyer, and Patrick Gallinari. Text classification : A sequential reading approach. In Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings, pages 411423, 2011.

[5] Gabriel Dulac-Arnold, Ludovic Denoyer, Philippe Preux, and Patrick Gallinari. Sequential approaches for learning datum-wise sparse representations. Machine Learning, 89(1-2):87/122, 2012.

[6] Gabriel Dulac-Arnold, Ludovic Denoyer, Nicolas Thome, Matthieu Cord, and Patrick Gallinari. Sequentially generated instance-dependent image representations for classification. International Conference on Learning Representations - ICLR 2014, 2014.

[7] Pallika H. Kanani and Andrew McCallum. Selecting actions for resource-bounded information extraction using reinforcement learning. In Proceedings of the Fifth International Conference on Web Search and Web Data Mining, WSDM 2012, Seattle, WA, USA, February 8-12, 2012, pages 253262, 2012.

[8] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. Recurrent models of visual attention. In Advances in Neural Information Processing Systems 27 : Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, pages 2204/2212, 2014.

[9] Zhixiang Eddie Xu, Matt J. Kusner, Kilian Q. Weinberger, Minmin Chen, and Olivier Chapelle. Classifier cascades and trees for minimizing feature evaluation cost. Journal of Machine Learning Research, 15(1):21132144, 2014.

International Collaborations

The topic of the thesis will allow interactions with different actors, in the context of national and european projects. Particularly, the student will develop collaborations with Ecole Centrale de Marseille, with IDIAP, with INRIA, and with different industrial partners (Google, Yahoo, ...) A joint proposal with Yahoo! (FREP project) has been made in May 2015 concerning this topic.

Relevance with the SMART Labex

The proposal is relevant with axis 3 "Humans at the convergence of digital and real environments" and particularly with sub-axis 3.1 "Large scale complex data processing" since it involves the development of sequential learning models for data processing and prediction. It is also marginally relevant with axis 2 "Interfaces and Interaction with humans". Indeed, the project also focuses on the problem of building predictive systems that can interact with humans. It thus involves the study of human-machine interactions from a pure machine learning point of view which is an original aspect of the proposal.