Sujet de stage: Interactive Machine Teaching for Social Agents

Contact :
Mohamed CHETOUANI
Institut des Systèmes Intelligents et de Robotique
CNRS UMR 7222
Sorbonne Université
Mohamed.Chetouani@sorbonne-universite.fr

Context:

In this work we focus on machine teaching [Zhu2018], which is the inverse problem of machine learning. Machine teaching studies the interaction between a teacher and a learner. Given a learning model and a target, the teacher aims to find an optimal set of training examples for the learner. In such a setting, the teacher selects a set of training examples based on previous learner performances. Machine teaching has been developed and applied in several contexts including education and adversarial settings (e.g., attacks).

This work is part of the European project ANIMATAS (www.animatas.eu) that aims to develop human-interaction computational models in the context of education. We target situations where humans have to learn from the interaction with artificial social agents.

Objectives

The main objective of this work is to develop machine-teaching algorithms for binary functions. In the context of education, this will translate to categorization tasks by binary properties: quadrilaterals, animals… These problems have addressed in supervised learning (see Zoo Animal Classification task in Kaggle).

Most of the machine-teaching approaches rely on the following phases: (i) the teacher observes the learner's current performance, (ii) then, the teacher provides the next teaching (training) examples, and (iii) the learner performs an update. Usually the learner is a machine-learning algorithm (e.g., linear classifier, SVM). This step is useful for the design of the machine-teaching models. In the our targeted applications, the learner is a human and the teacher is a machine able to provide relevant examples to the human learner. Several assumptions could be made on human teaching including "back-box" [Dasgupta2019]. In this work, we target learners that make mistakes or either don't provide responses to the teacher. We will firstly simulate these features then we will evaluate it with real humans. The teacher could also rely on social interactions. In particular, we aim to introduce queries that will result in explicitly asking the human to perform classification.

The main steps are:

- Development of an educational task using binary properties.
- Development a first interactive machine-teaching algorithm (omniscient, black-box) using linear classifier (teacher)
- Analysis of performance of the teacher-learner interactions
- Modification of the machine-teaching model by introducing: (i) learner: mistakes, sparse answers and (ii) teacher: query teacher for classification.
- If possible, evaluation with real humans.

[Zhu2018] X. Zhu, A. Singla, S. Zilles, A. N. Rafferty (2018) An Overview of Machine Teaching; ArXiv 1801.05927

[Dasgupta2019] S. Dasgupta, D. Hsu, S. Poulis, X. Zhu (2019). Teaching a black-box learner. In The 36th International Conference on Machine Learning (ICML).

Skills: Python, Machine learning

Duration: 5-6 months