CSC one-pager round 2023

Project title: Explainable and Reliable Transfer Learning

Principal Investigators:

Dr. E.N. Smirnov, Assistant Professor, Department of Advanced Computing Sciences, Maastricht University

Dr. K. Driessens, Associate Professor, Department of Advanced Computing Sciences, Maastricht University

Promotor: Prof. Dr. Gerhard Weiss, Department of Advanced Computing Sciences, Maastricht University

Proposal (250 words):

Introduction: Transfer learning aims at improving predictive models for a target domain by exploiting predictive models and data from related source domains. While there are many successful applications, the mechanics of transfer learning are not always well understood and, thus, there is neither a guarantee that a particular task will be solved nor a clear view on how the transfer is accomplished. This assumes that in practice transfer learning needs to be enhanced by *reliability* and *explainability*. *Reliable* transfer learning could give a user the assurance that the probability of failure is bounded by a pre-specified significance level. *Explainable* transfer learning could make it clear to a user what has been transferred and why, or what are reasons for successful transfer or transfer failure. Successful development of Explainable and Reliable Transfer Learning will create a kind of trust from a user perspective that can be a game changer for applied transfer learning.

Objectives: This proposal aims at developing an approach to transfer learning that provides statistical reliability and explanations for failure/success in a predefined language.

Methods: Techniques from reliable machine learning and explanation-based learning will be used. More precisely, to introduce reliability in transfer learning we will employ methods from conformal prediction. To introduce explainability we will source from relational learning in addition to explanation-based learning.

Impact: The proposed approach will be applied for data fusion problems for a multi-national company.

Team: The supervisors' team has experience in theory and practice of transfer learning as well as in guiding PhD students on this topic.

Requirements candidate: Solid background in math, computer science, and machine learning on a Master level. Good English language skills.

Keywords: machine learning, transfer learning, reliability, explainability, conformal prediction

Top 5 relevant selected publications:

1. A Kroner, M Senden, K Driessens, R Goebel, *Contextual encoder-decoder network for visual saliency prediction*, Neural Networks 129, 261-270, 2020
2. S Zhou, E Smirnov, G Schoenmakers, R Peeters, X Wu, *Conformal Feature-Selection Wrappers and ensembles for negative-transfer avoidance*, Neurocomputing 397, 309-319, 2020
3. S Zhou, EN Smirnov, R Peeters, *Conformal region classification with instance-transfer boosting,* International Journal on Artificial Intelligence Tools 24(6), 1-25, 2015
4. H Bou Ammar, D Mocanu, M Taylor, K Driessens, K Tuyls, G Weiss, *Automatically mapped transfer between reinforcement learning tasks via three-way restricted Boltzmann machines*, In Proceeding of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD-2014), Springer, 449-464, 2014
5. J Ramon, K Driessens, T Croonenborghs, *Transfer learning in reinforcement learning problems through partial policy recycling*, In Proceeding of the European Conference on Machine Learning (ECML-2007), Springer, 699-707, 2007