

The Research Group
Artificial Intelligence Lab

has the honor to invite you to the public defense of the PhD thesis of

Isel del Carmen Grau Garcia

to obtain the degree of Doctor of Sciences

Title of the PhD thesis:

Self-labeling Grey-box Model: An Interpretable Semi-supervised Classifier

Promotor:

Prof. dr. Ann Nowé

Co-promotor:

Prof. dr. Maria Matilde Garcia Lorenzo

The defense will take place on

Friday, October 9, 2020 at 17:00h.

The defense can be followed through a live stream. An online link for live-streaming will be soon added here: <https://bit.ly/36t0fxD>

Members of the jury

Prof. dr. Beat Signer (VUB, chair)

Prof. dr. Wim Vranken (VUB, secretary)

Prof. dr. Tom Lenaerts (VUB)

Prof. dr. Sonia Van Dooren (VUB)

Prof. dr. Koen Vanhoof (UHasselt)

Prof. dr. Chris Cornelis (UGent)

Curriculum vitae

Isel Grau graduated as a MSc in Computer Science from the Central University of Las Villas (Cuba) in 2014 and started her PhD studies at the Artificial Intelligence Lab of the VUB in 2015. Her PhD research focusses on interpretable machine learning and semi-supervised classification techniques with applications to bioinformatics problems. She has (co-)authored 14 peer-reviewed papers in international conferences and journals during her doctoral research. She has presented her research at several international events.

Abstract of the PhD research

In the context of some machine learning applications, obtaining data points is a relatively simple process, yet labeling them could become quite expensive or tedious. Such scenarios lead to datasets with few labeled points and a higher number of unlabeled ones. Semi-supervised classification techniques combine labeled and unlabeled data during the learning process in order to improve baseline supervised methods that use only labeled data. Unfortunately, most successful semi-supervised classifiers are complex structures that do not allow explaining their predictions, thus behaving like black boxes. However, there is an increasing number of problem domains in which experts demand a clear understanding of the decision process. Intrinsically interpretable classifiers (i.e., white-box models) are transparent structures that allow performing predictions, obtaining an associated explanation, and inspecting the model as a whole. Nevertheless, these advantages generally come at the cost of performance in terms of accuracy.

In this thesis, we propose the self-labeling grey-box model, a semi-supervised classifier aiming at providing a suitable balance between accuracy and interpretability. The self-labeling grey-box uses an accurate black-box classifier for labeling the unlabeled data and a white-box surrogate classifier for building an interpretable model. Since the self-labeling process can propagate errors, we propose two amending procedures based on class membership probabilities and certainty measures coming from the field of rough sets theory. The experimental study shows the influence of increasing ratios of labeled and unlabeled data across benchmark datasets. Moreover, we study the effect of different black-box, white-box base classifiers, as well as the two proposed amending procedures in terms of both accuracy and interpretability. The results support the interpretability of our classifier using simplicity and transparency as proxies while attaining superior prediction rates when compared with state-of-the-art self-labeling classifiers. Additionally, we illustrate the applicability of the self-labeling grey-box classifier with preliminary results in two case studies from the field of bioinformatics. The first task concerns the detection of disease-causing genomic variants in a rare disease, while the second application tackles the prediction of early folding in proteins. Both case studies require an interpretable model able to leverage extra unlabeled data.